

-2251-

Query: 1 MYLIEPIRNGKRITDGAIALAMQVYILQNVFLDDDDILFPYYCDPKVEIGKFQNAVETNQ 60
 MYLIEPIRNGKRITDGA+ALAMQVY+ +N+FLDDDDILFPYYCDPKVEIGKFQNAV+ETNQ
 Sbjct: 1 MYLIEPIRNGKRITDGAVALAMQVYVQENLFLDDDDILFPYYCDPKVEIGKFQNAVETNQ 60

Query: 61 EYLKEHDIPVVRDRTGGGAVYVDSGAVNICYLMKDHGQFGDFKRAYEPAIKALKTIGASS 120
 EYLKEH IPVVRDRTGGGAVYVDSGAVNICYL+ D+G FGDFKR Y+PAI+AL LGA+
 Sbjct: 61 EYLKEHHIPVVRDRTGGGAVYVDSGAVNICYLINDNGIFGDFKRTYQPAIEALHHLGATE 120

Query: 121 VEMRERNDLVIDGKKVSGAAMTIVNGRIYGGYSLLLDVFDFAMEKVLNPNRKKIESKGIK 180
 VEM RNDLVIDGKKVSGAAMTI NGR+YGGYSLLLDVFDF+AMEK L PNRKKIESKGI+
 Sbjct: 121 VEMSGRNDLVIDGKKVSGAAMTIANGRVYGGYSLLLDVFDFEAMEKALKPNRKKIESKGIK 180

Query: 181 SVRSRVGDIRSHLSEDIRHTITDQFKDLMVCQLLHIDHIDQAKRYHLTEKDWAAIDALAD 240
 SVRSRVG+IR HL+ Y+ IT ++FKDLMVCQLL I+ I QAKRY LTEKDW IDAL +
 Sbjct: 181 SVRSRVGNIREHLAPQYQGITEEFKDLMLVCQLLQIETISQAKRYDLTEKDWQIDALTE 240

Query: 241 EKYKNWDWNYGNSPQYSYHRDARFSPGTYDFHLEIEKGIIITNCRIYGDFFSSKDISDIEN 300
 KY NW+WNYGN+PQY YHRD RF GT D HL+I+KG I CRIYGDFF DI+++E
 Sbjct: 241 RKYHNWEWNYGNAPQYRYHRDGRFTGTVDIHLDIKKGYIAACRIYGDFFGKADIAELEG 300

Query: 301 LLIGCPMKEELVLEKLSTLSLEDYFGQTSPEEIKAVLFS 339
 LIG M++E VL L+ + L Y G + EE+ ++FS
 Sbjct: 301 HLIGTRMEKEDVLATLNAIDLAPYLGAITAEELGDLIFS 339

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2004

A DNA sequence (GBSx2114) was identified in *S. agalactiae* <SEQ ID 6197> which encodes the amino acid sequence <SEQ ID 6198>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 196 - 212 (196 - 212)

----- Final Results -----
 bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB49329 GB:U39612 formyl-tetrahydrofolate synthetase
 [Streptococcus mutans]
 Identities = 432/556 (77%), Positives = 493/556 (87%)

Query: 1 MKTDIEIAQSVALKPPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGKLILVT 60
 MKTDIEIAQSV L+PI +V+++GI FDD+ELYGKYKAKL+FDKI+AV+ GKL+LVT
 Sbjct: 1 MKTDIEIAQSVDLRPIITNVVKLGIDFDDLELYGKYKAKLTFDKIKAVEENAPGKLVLVT 60

Query: 61 AINPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVGMIGGAAGGGYAQVLPME 120
 AINPTPAGEGKST++IGLADALNKIGKKTMI+REPSLGPVGMIGGAAGGGYAQVLPME
 Sbjct: 61 AINPTPAGEGKSTITIGLADALNKIGKKTMIAREPSLGPVGMIGGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALLDNHIIHQGNELDIDQRRVIWKRVDLNDRLRQVIVG 180
 DINLHFTGDMHAITTANNALSAL+DNH+HQGNEL IDQRR+IWKRVVDLNDRLR V VG
 Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLHQGNELGIDQRRIIWKRVVDLNDRLRHVTVG 180

Query: 181 LGSPVNGIPREDGFDTIVASEIMAILCLATDLSDLKKRLSNIVVAYSRRNRKPIYVKDLKI 240
 LGSP+NGIPREDGFDTIVASEIMAILCLAT++ DLK+RL+NIV+ Y +R P+YV+DL++
 Sbjct: 181 LGSPINGIPREDGFDTIVASEIMAILCLATNVEDLKERLANIVIGYRFRSPVYVRDLEV 240

Query: 241 EGALTLILKDTIKPNLVQTIYGTTPALVHGGPFANIAHGCNSVLATSTALRLADYVTEAG 300
 +GAL LILK+ IKPNLVQTIYGTTPA VHGGPFANIAHGCNSVLATSTALRLADY +TEAG

-2252-

Sbjct: 241 QGALALILKEAIKPNLVQTIYGTTPAFVHGGPFANIAHGCNSVLATSTALRLADYTITEAG 300

Query: 301 FGADLGAEKFLDIKTPNLPTSPDAIVIVATLRALKMHGGVSKEDLSQENVEAVKRGFTNL 360
FGADLGAEKFLDIK PNLPTSPDA+VIVAT+RALKM+GGV+K+ L+QENVEAVK GF NL

5 Sbjct: 301 FGADLGAEKFLDIKAPNLPTSPDAVVIVATIRALKMNGGVAKDALNQENVEAVKAGFANL 360

Query: 361 ERHVNMMRQYGVPPVVAINQFTADTESEIATLKTLCNIDVAVELASVWEDGADGGLELA 420
RHV NMR+YGVPPVVAIN+F DT EIA L+ LC+ IDV VELASVW +GADGG++LA

10 Sbjct: 361 ARHVENMRKYGVPPVVAINFEITDTNDEIAVLRLNCAIDVPVELASVWANGADGGVDLA 420

Query: 421 QTVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSDNQWD 480
T+ N IE S+YKRLY++ ++EEK+ +I +IY +KV F KA+ Q+ + NGWD

15 Sbjct: 421 NTLINTIENNP SHYKRLYDNNLSVEEKVTEIAKEIYRADKVI FEKKAKTQIAQIVKNGWD 480

Query: 481 KMPICMAKTQYSFSDNPNLLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPA 540
+PICMAKTQYSFSD+P LLGAPT FDIT+RE VPK GAGFIVALTGDV+TMPGLPKKPA

20 Sbjct: 481 NLPICMAKTQYSFSDDPKLLGAPTGFDTITRELVPKLGAGFIVALTGDVMTMPGLPKKPA 540

Query: 541 ALNMDVLEDGTAGLGF 556
ALNMDV DGTAGLGF

20 Sbjct: 541 ALNMDVAADGTALGLF 556

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6199> which encodes the amino acid sequence <SEQ ID 6200>. Analysis of this protein sequence reveals the following:

25 Possible site: 50
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.49 Transmembrane 196 - 212 (196 - 212)

30 ----- Final Results -----
bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP:AAB49329 GB:U39612 formyl-tetrahydrofolate synthetase
[Streptococcus mutans]
Identities = 432/556 (77%), Positives = 490/556 (87%)

40 Query: 1 MKSDIEIAQSVLPITDIVKKVGIDDDIELYGYKAKLSFEKMKAVEANEPGKLILVT 60
MK+DIEIAQSV L+PIT++VKK+GID DD+ELYGYKAKL+F+K+KAVE N PGKL+LVT

Sbjct: 1 MKTDIEIAQSVDLRPITNVKKLGIDFDDLELYGYKAKLTFDKIKAVEENAPGKLVLVT 60

Query: 61 AINPTPAGEGKSTMSIGLADALNQMGKKTMLALREPSLGPVPMGIKGAAGGGYAQVLPME 120
AINPTPAGEGKST++IGLADALN++GKKTMA+REPSLGPVPMGIKGAAGGGYAQVLPME

45 Sbjct: 61 AINPTPAGEGKSTITIGLADALNKIGKKTMAIREPSLGPVPMGIKGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALIDNHLQQGNDLGIDPRRIIWKRVLNDRALRQVIVG 180
DINLHFTGDMHAITTANNALSALIDNHL QGN+LGID RRIIWKRVLNDRALR V VG

50 Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLHQGNELGIDQRRRIIWKRVLNDRALRHVTVG 180

Query: 181 LGSPVNGVPREDGFDITVASEIMAILCLATDLKDLKKRLADIVVAYTYDRKPVYVRDLKV 240
LGSP+NG+PREDGFDITVASEIMAILCLAT++DLK+RLA+IV+ Y +DR PVYVRDL+V

55 Sbjct: 181 LGSPINGIPREDGFDITVASEIMAILCLATNVEDLKERLANIVIGYRFRDRSPVYVRDLEV 240

Query: 241 EGALTILKDAIKPNLVQTIYGTTPALIHGGPFANIAHGCNSVLATSTALRLADYTITEAG 300
+GAL LILK+AIKPNLVQTIYGTTPA +HGGPFANIAHGCNSVLATSTALRLADYT+TEAG

60 Sbjct: 241 QGALALILKEAIKPNLVQTIYGTTPAFVHGGPFANIAHGCNSVLATSTALRLADYTITEAG 300

Query: 301 FGADLGAEKFLNIKVPNLKPAPDAIVIVATLRALKMHGGVAKSDLAANCEAVRLGFANL 360
FGADLGAEKFL+IK PNL+PDA+VIVAT+RALKM+GGVAK L EN EAV+ GFANL

60 Sbjct: 301 FGADLGAEKFLDIKAPNLPTSPDAVVIVATIRALKMNGGVAKDALNQENVEAVKAGFANL 360

Query: 361 KRHVENMRQFKVPVVVAINEFVADTEAEIATLKALCEEIKVPVELASVWANGAEGGLALA 420
RHVENMR++ VPVVVAINEF+ DT EIA L+ LC I VPVELASVWANGA+GG+ LA

65 Sbjct: 361 ARHVENMRKYGVPPVVVAINEFITDTNDEIAVLRLNCAIDVPVELASVWANGADGGVDLA 420

-2253-

Query: 421 KTVVRVIDQEADYKRLYSDDEDTLEEKVINIVTQIYGGKAVQFGPKAKTQLKQFAEFGWD 480
 T++ I+ + YKRLY + ++EEKV I +IY V F KAKTQ+ Q + GWD
 Sbjct: 421 NTLINTIENNPISHYKRLYDNNLSVEEKVTEIAKEIYRADKVFIEKKAKTQIAQIVKNGWD 480

Query: 481 KLPVCMAKTQYSFSDNPSSLGAPTDFDITIREFVPKTGAGFIVGLTGDVMTMPGLPKVPA 540
 LP+CMAKTQYSFSDNP LLGAPT FDIITIRE VPK GAGFIV LTGDVMTMPGLPK PA
 Sbjct: 481 NLPICMAKTQYSFSDDPKLLGAPTDFDITIRELVPKLGAGFIVALTGDVMTMPGLPKKPA 540

Query: 541 AMAMDVAENGTALGLF 556
 A+ MDVA +GTALGLF
 Sbjct: 541 ALNMDVAADGTALGLF 556

An alignment of the GAS and GBS proteins is shown below.

Identities = 452/556 (81%), Positives = 513/556 (91%)

Query: 1 MKTDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGKLILVT 60
 MK+DIEIAQSVAL+PI +IV++VGI DDIELYGKYKAKLSF+K++AV++ + GKLILVT
 Sbjct: 1 MKSDIEIAQSVALQPIITDIVKKVGIDGDDIELYGKYKAKLSFEKMKAVEANEPGKLILVT 60

Query: 61 AINPTPAGEGKSTMSIGLADALNKIGKKTMLALREPSLGPVGMIGGAAGGGYAQVLPME 120
 AINPTPAGEGKSTMSIGLADALN++GKKTMLALREPSLGPVGMIGGAAGGGYAQVLPME
 Sbjct: 61 AINPTPAGEGKSTMSIGLADALNQMGKKTMLALREPSLGPVGMIGGAAGGGYAQVLPME 120

Query: 121 DINLHFTGDMHAITTANNALSALLDNHIIHQGNELDIDQRRVIWKRVDLNDRALRQVIVG 180
 DINLHFTGDMHAITTANNALSAL+DNH+ QGN+L ID RR+IWKRV+DLNDRALRQVIVG
 Sbjct: 121 DINLHFTGDMHAITTANNALSALIDNHLQGGNDLGIDPRRIIWKRVLDLNDRALRQVIVG 180

Query: 181 LGSPVNGIPREDGFDTIVASEIMAILCLATDLSDLKKRLSNIVVAYSRRNKPIYVKDLKI 240
 LGSPVNG+PREDEGFDITVASEIMAILCLATDL DLKKRL++IVVAY+ +RKP+YV+DLK+
 Sbjct: 181 LGSPVNGVPREDEGFDITVASEIMAILCLATDLKDLKKRLADIVVAYTYDRKPVYVRDLKV 240

Query: 241 EGALTILIKDITIKPNLVQTIYGPALVHGGPFANIAHGCNSVLATSTALRLADYVVTAG 300
 EGALTILIKD IKPNLVQTIYGPAL+HGGPFANIAHGCNSVLATSTALRLADY VTEAG
 Sbjct: 241 EGALTILIKDAIKPNLVQTIYGPALIHGGPFANIAHGCNSVLATSTALRLADYTVTEAG 300

Query: 301 FGADLGAEKFLDIKTPNLPTSPDAIVIVATLRALKMHGGVSKEDLSQENVEAVKRGFTNL 360
 FGADLGAEKFL+IK PNLPT+PDAIVIVATLRALKMHGGV+K DL+ EN EAV+ GF NL
 Sbjct: 301 FGADLGAEKFLNIKVPNLKPAPDAIVIVATLRALKMHGGVAKSDLAENCEAVRLGFANL 360

Query: 361 ERHVNNMRQYGVVPPVAINQFTADTESEIATLKLCSNIDVAVELASVWEDGADGGLELA 420
 +RHV NMRQ+ VPVPPVAIN+F ADTE+EIATLK LC I V VELASVW +GA+GGL LA
 Sbjct: 361 KRHVNNMRQYGVVPPVAINQFTADTESEIATLKLCSNIDVAVELASVWANGAEGGLALA 420

Query: 421 QTVANVIETQSSNYKRLYNDEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSNDNGWD 480
 +TV VI+ ++++YKRLY+DEDT+EEK+ IVT+IYGG V FGPKA+ QLK+F++ GWD
 Sbjct: 421 KTVVRVIDQEADYKRLYSDDEDTLEEKVINIVTQIYGGKAVQFGPKAKTQLKQFAEFGWD 480

Query: 481 KMPICMAKTQYSFSDNPSSLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPA 540
 K+P+CMAKTQYSFSDNP+LLGAPTDFDIT+REFVPKTGAGFIV LTGDV+TMPGLPK PA
 Sbjct: 481 KLPVCMAKTQYSFSDNPSSLGAPTDFDITIREFVPKTGAGFIVGLTGDVMTMPGLPKVPA 540

Query: 541 ALNMDVLEEDGTAIGLF 556
 A+ MDV E+GTA+GLF
 Sbjct: 541 AMAMDVAENGTALGLF 556

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9057> which encodes amino acid sequence <SEQ ID 9058>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.49 Transmembrane 516 - 532 (516 - 533)

----- Final Results -----
 bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2254-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 604 bits (1540), Expect = e-174
 Identities = 304/555 (54%), Positives = 389/555 (69%), Gaps = 2/555 (0%)

Query: 4 SDIEIANSVTMEPIISKVADQLGIDKEALCLYGKYKAKIDARQLVALKNKPDGKLILVTAI 63
 +DIEIA SV ++PI+++ +Q+GI + + LYGKYKAK+ ++ A+K++ GKLILVTAI
 Sbjct: 3 TDIEIAQSVALKPIAEIVEQVGIGFDDIELYGKYKAKLSFDKIEAVKSQKVGKLILVTAI 62

Query: 64 SPTPAGEGKTTTSVGLVDALSAIGKKAVIALREPSLXXXXXXXXXXXXXXXXXXPMEDI 123
 +PTPAGEGK+T S+GL DAL+ IGKK +IALREPSL PMEDI
 Sbjct: 63 NPTPAGEGKSTMSIGLADALNKIGKKTMIALREPSLGPVMGIKGAAGGGYAQVLPMEI 122

Query: 124 NLHFTGDFHAIGVANNLLAALIDNHIHGNLSLGIDSRITWKRVDMDNRQLRHIVDGLQ 183
 NLHFTGD HAI ANN L+AL+DNHIH GN L ID RR+ WKRVD+NDR LR ++ GL
 Sbjct: 123 NLHFTGDMHAITTANNALSALLDNHIHQGNELDIDQRRVIWKRVDLNDRLRQVIVGLG 182

Query: 184 GKVNGIPREDGYDITVASEIMAILCLSENISDLKARLEKIIIGYNYQGEPTVXXXXXXXX 243
 VNGIPREDG+DITVASEIMAILCL+ ++SDLK RL I++ Y+ +P+
 Sbjct: 183 SPVNGIPREDGFDITVASEIMAILCLATDLSDLKKRLSNIVVAYSRRNKPIYVKDLKIEG 242

Query: 244 XXXXXXXXXIHPNLVQTLHTPALIHGGPFANIAHGCNSVLATKLALKYGDYAVTEAGFG 303
 I PNLVQT+ TPAL+HGGPFANIAHGCNSVLAT AL+ DY VTEAGFG
 Sbjct: 243 ALTLILKDTIKPNLVQTIYGTALVHGGPFANIAHGCNSVLATSTALRLADYVTEAGFG 302

Query: 304 ADLGAEKFDIDIKRMSGLRPAAVLVATIRALKMHGGVPKADLATENVQAVVDGLPNLKD 363
 ADLGAEK+DIK P A+V+VAT+RALKMHGGV K DL+ ENV+AV G NL++
 Sbjct: 303 ADLGAEKFLDIKTPNLPTSPDAIVIVATLRLKMHGGVSKEDLSQENVEAVKRGFTNLER 362

Query: 364 HLANIQDVYGLPVVVAINKFPLDTEALQAVYDACDKRGVDVVISDVWANGGAGGRELA 423
 H+ N++ YG+PVVVAIN+F DT++E+ + C V V ++ VW +G GG ELA+
 Sbjct: 363 HVNNMRQ-YGVPVVVAINQFTADTESEIATLTKLCSNIDVAVELASVWEDGADGGLELAQ 421

Query: 424 KVVTLAE-QDNQFRFVYEEDDSIETKLTIVTKVYGGKGINLSSAAKRELADLERLGFGN 482
 V + E Q + ++ +Y ++D+IE K+ KIVTK+YGG ++ A+ +L + G+
 Sbjct: 422 TVANVIETQSSNYKRLYNEDTIEEKIKKIVTKIYGGNKVHFGPKAQIQLKEFSNGWDK 481

Query: 483 YPICMAKTQYSFSDDAKKLGAPTDFVTISNLKVSAGAGFIVALTGAIMTMPGLPKVPAS 542
 PICMAKTQYSFSD+ LGAPTDF +T+ GAGFIVALTG ++TMPGLPK PA+
 Sbjct: 482 MPICMAKTQYSFSDNPNLGAPTDFDITVREFVPKTGAGFIVALTGDVLTMPGLPKKPAA 541

Query: 543 ETIDIDEEGNITGLF 557
 +D+ E+G GLF
 Sbjct: 542 LNMDVLEDGTAIGLF 556

SEQ ID 6198 (GBS131) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 6; MW 64.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 4; MW 90kDa).

GBS131-GST was purified as shown in Figure 201, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2005

A DNA sequence (GBSx2115) was identified in *S.agalactiae* <SEQ ID 6201> which encodes the amino acid sequence <SEQ ID 6202>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.03 Transmembrane 34 - 50 (29 - 56)

-2255-

INTEGRAL Likelihood = -7.70 Transmembrane 90 - 106 (84 - 110)
 INTEGRAL Likelihood = -1.97 Transmembrane 62 - 78 (62 - 78)
 INTEGRAL Likelihood = -0.69 Transmembrane 275 - 291 (275 - 291)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA88609 GB:M37842 unknown protein [Streptococcus mutans]

Identities = 243/373 (65%), Positives = 302/373 (80%), Gaps = 1/373 (0%)

15 Query: 71 IGAVLYLVNSEMDALSRVTWLILVMIAPLLGAMFLMYTKFDWGYRGLKQRLLETILIDESQI 130
 IG+VLYLVNS+MD LS +TWL++++ P+LG +FL+YTK DWGYR LK ++ +
 Sbjct: 2 IGSVLYLVNSQMDTSLIITWLLVILFPFIPILGTLFLIYTKQDWGYRELKSLIKKSTQAIKP 61

Query: 131 YLEDDPETLNQLKSSTSTTYHLVQYFEKAHGNFPVYRNTDVTFLPTGEAFFEKMKEELLK 190
 Y + D L +LK S + TY+L QY ++ G FPVY+NT VT+ P G++ FE+MK++LLK
 20 Sbjct: 62 YFQYDQRILYKLKESHARTYNLAQYLHRS-GGFPVYKNTKVITYFPNGQSKFEEMKKQLLK 120

Query: 191 AKKYIFLEFFIIDEGIMWGEILSILEQKVEEGVEVRILYDGMIEITKLSFDYTKRLEKIG 250
 A+K+IFLE+FII EG+MWGEILSILEQKV+EGVEVR++YDGM+E++ LSFDY KRLEKIG
 25 Sbjct: 121 AEKFIFLEYFIIAEGLMWGEILSILEQKVEGVEVRVMYDGMLELSTLSFDYAKRLEKIG 180

Query: 251 IKAKAFSPISPFISTYYNYRDHRKIVVIDGVVGMTGGVNLADHEYINHIEFGHWKDSGIM 310
 IKAK FSPI+PF+STYYNYRDHRKI+VID V GG+NLADHEYIN IE FG+WKD+ +M
 Sbjct: 181 IKAKVFSPIPFVSTYYNYRDHRKILVIDNKAFAFGGINLADEYINQIERFGYWKDTAVM 240

30 Query: 311 LKGKAVDSFLLLFLQMWSSITEKMLVAPYLGVDHDLVENEGYVIPYGDSPDLTDKVGENV 370
 L+G+ V SF L+FLQMWSS T + APYL + + GYVIPY DSPLD +KVGENV
 Sbjct: 241 LEGEGVASFTLMFLQMWSTTNKDYEFAPYLTQNFHEIVANGYVIPYSDSPDLDEKVGENV 300

Query: 371 YIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRIIMPGIPDKPIPYALAKTY 430
 YIDILN AR+YVYIMTPYLILDSE+EHA+QFAAERGVDV+IIMPGIPDK +P+ALAK Y+
 35 Sbjct: 301 YIDILNQARDYVYIMTPYLILDSEMEHALQFAAERGVDVKIIMPGIPDKKVPFALAKRYF 360

Query: 431 QALTSGVKIYEF 443

AL +GVKIYE+

40 Sbjct: 361 PALLDAGVKIYEF 373

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6203> which encodes the amino acid sequence <SEQ ID 6204>. Analysis of this protein sequence reveals the following:

Possible site: 47

45 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -8.86 Transmembrane 84 - 100 (81 - 104)
 INTEGRAL Likelihood = -8.33 Transmembrane 28 - 44 (23 - 49)
 INTEGRAL Likelihood = -6.74 Transmembrane 56 - 72 (53 - 74)

50 ----- Final Results -----

bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

>GP:AAA23240 GB:J02911 formyltetrahydrofolate synthetase (FTHFS)

(ttg start codon) (EC 6.3.4.3) [Moorella thermoacetica]

Identities = 350/557 (62%), Positives = 438/557 (77%), Gaps = 2/557 (0%)

60 Query: 2 VLSDIEIANSVTMEPIISKVADQLGIDKEALCLYGKYKAKIDARQLVALKNKPDGKLILVT 61
 V SDIEIA + M+P+ ++A LGI ++ + LYGKYKAKI LK+KPDGKLILVT
 Sbjct: 4 VPSDIEIAQAAMKPMVMEIARGLGIQEDEVLYGKYKAKISLDVYRRLKDKPDGKLILVT 63

Query: 62 AISPTPAGEGKTTTSGVLVDALSAIGKKAVIALREPSLGPVFGVKGAAGGGAQVVPME 121

5
10
15
20
25
30
35
40
45
50
55
60
65

An alignment of the GAS and GBS proteins is shown below.

-2257-

Identities = 362/524 (69%), Positives = 437/524 (83%)

5 Query: 8 LISNKVKIVRLNKS KSKSLRGIFSRRTTVIAILLQLLFLLASYSWLEQYRVWLATVEH 67
+I K K+ LL+K K LRGIFSRRTT+I +L+ILQL+FL SY+W+EQYRVW+ +E
Sbjct: 2 IIKKKAKVKYLLHKHKGHFLRGIFSRRTTIVLLIILQLVFLFQSYAWMEQYRVWITILES 61

10 Query: 68 ILTIGAVLYLVNSEMDALSRVTWLILVMIAPLLGAMFLMYTKFDWGYRGLKQRLETLIDE 127
+ I VLYLVNS+MDA+SR+TWLIL+MIAPLLG++FL+YTK DWGYRGLKQR+ L+D
Sbjct: 62 VFAITIVLYLVNSDMDAISRTWLILIMIAPLLGSLFLIYTKLDWGYRGLKQRINHLVLDL 121

15 Query: 128 SQIYLEDDPETLNQLKSSTSTTYHLVQYFEKAHGNFPVYRNTDVTFLPTGEAFFEKMKEE 187
S YL DD L LK STSTTYHLVQY E++ GNFP+Y NT VT+ PTGE FF+ +KE+
Sbjct: 122 SAPYLSDDDAILEVLKDDSTSTTYHLVQYLERSGNFPFIYNNTRVTFPTGETFFDSLKEQ 181

20 Query: 188 LLKAKKYIFLEFFIIDEIGIMWGEILSILEQKVEEGVEVRILYDGMIEITKLSFDYTKRLE 247
L AKKYIFLEFFII EG MWGEILSILE+KV EGVEVR+L+DGM E++ LS DY KRLE
Sbjct: 182 LFLAKKYIFLEFFIIAEGQMWGEILSILEKKVSEGVEVRVLFDMNLSLSSDYAKRLE 241

25 Query: 248 KIGIKAKAFSPISPFISTYNYNRDHRKIVVIDGVGMTGGVNLAD EYINHIELFGHWKDS 307
+IGIKAK+F PISPFISTYNYNRDHRKIVVIDG V TGG+NLAD EYIN +E FGHWKD+
Sbjct: 242 QIGIKAKSFLPISPFISTYNYNRDHRKIVVIDGEVSFTGGINLAD EYINEVERFGHWKDA 301

30 Query: 308 GIMLKGKAVDSFLLFLQMW SITEEKMVLVAPYLG VHDLDVENEGYVIPYGDSPLDTDKVG 367
G+ML+G+A DSFL+LFLQMW SITE+++++ PYL H + ++GYVIPYGDSPLDTDK+G
Sbjct: 302 GLMLEGEATDSFLILFLQMW SITEKELIIDPYLSDHSLKLPDGVIPYGDSPLDTDKIG 361

35 Query: 368 ENVYIDILNHAREYVYIMTPYLILDSELEHAIQFAAERGVDVRIIMPGIPDKPIPYALAK 427
+NVYIDILNHA+EYVYIMTPYLILDSE+EHA++FA+ERGVD+RIIMPG+PDK +PYALAK
Sbjct: 362 KNVYIDILNHAKEYVYIMTPYLILDSEMEHALRFASERGVDIRIIMPGVPDKGV PYALAK 421

40 Query: 428 TYYQALTKSGVKIYEYTLGFVHSKI FLSDNTKAVVGTINLDYRSLYHHFECAVLYKYKDA 487
TYY+AL SGVKIYEY GFVHSK+F+SDNTKAVVGTINLDYRSLYHHFECA YLY+V
Sbjct: 422 TYYKALMSSGVKIYEYQPGFVHSKVFISDNTKAVVGTINLDYRSLYHHFECA TYLYRVSV 481

45 Query: 488 IQDIYRDYMDTLNKSRLVSLKDINNIPKFQKVIGIVTKTIAPLL 531
I DI D+ + +S L++ + P +QK+IG++ + IAPLL
Sbjct: 482 IADIVNDFNEAQKQSLMTSDHLTQRPWYQKLIGLLVRIIAPLL 525

A related GBS gene <SEQ ID 8953> and protein <SEQ ID 8954> were also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 6
McG: Discrim Score: -8.80
GvH: Signal Score (-7.5): -1.94
Possible site: 53

>>> Seems to have no N-terminal signal sequence
ALOM program count: 4 value: -10.03 threshold: 0.0
INTEGRAL Likelihood = -10.03 Transmembrane 34 - 50 (29 - 56)
INTEGRAL Likelihood = -7.70 Transmembrane 90 - 106 (84 - 110)
INTEGRAL Likelihood = -1.97 Transmembrane 62 - 78 (62 - 78)
PERIPHERAL Likelihood = 1.22 199
modified ALOM score: 2.51

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

32.5/57.2% over 498aa

Bacillus firmus
SP|O66043| CARDIOLIPIN SYNTHETASE (EC 2.7.8.-) (CARDIOLIPIN SYNTHASE) (CL SYNTHASE).
Insert characterized
GP|2952028|gb|AAC05444.1||U88888 cardiolipin synthase Insert characterized

ORF01572 (409 - 1893 of 2193)

%Match = 17.9

```
%Identity = 32.5  %Similarity = 57.1
```

Matches = 162 Mismatches = 204 Conservative Sub.s = 123

153 183 213 243 273 303 333 363
NLQLSIWMWF*KTVOPLDYFK**RGRACDASLFLLGIRF*LEII*NNRMLFK*QYAIIK*LIIWRGEKLISNKVKIVRLNK

393 423 447 477 507 528 558 588
SKKSLRGIGFSRTTIVIAILLILOLFF--LLASYSWLEQYRVMLATVEHIILT---IGAVILYLVNSEMDALSRVTWLILVMI

MKNRLNVLAFALLFAALYISRGFLQSMVGTLSVVFITLSVIFIGIIFFEN--RHPTKTLTWLLVLAA

618 648 678 705 735 765 789 819
APLLGAMFLMYTKFDWGYRGLKORLETLIDESQYLE-DDPETINOLKSSTSTTYHLVOYFEKAH--GNPPVYRNTDVT

FPVVG--FFFYLMFGQNHRKSKRFSKKAIEDERAFQKIEGQRQLNE-EQLKKMGGHQQLFLRAHKLGNPISFSSETKV

849 879 909 939 969 999 1029 1059
LPTGEAFFEFKMKELLKAKKYIFLEFFFIIDEGIMWGEILSILEOKVEEGVEVRILYDGMIEITKLSFDYTKRLEKIGIKA

LTGDKETYAHILQALKMAEHHIHLEYYIVRHDDLGNQIKDILISKAKEGVHVRFLYDG-VGSWKLKSKSYVEELRDAGVEM

1086 1116 1146 1176 1206 1236 1266 1293
KAFSPIS-PFISTYYNRYDRHKIVVIDGVGMTGGVNLAD EYINHIELFGHWKDSGIMLGKGAVD SFLLLFLQWSTI-TE

VSFSPVKLPFLTHTTINYRNHRKIIVIDGVVGFVGGNLIGDEYLKGDAYFGYWRDTHLYVRGEAVRTLQLIFLQDWHYQTG

1323 1353 1383 1413 1443 1473 1503 1533
EKMLVAPYLGVHDDLVENEGYVIPYGDSPLDTDKGVENVYIDILNHAREYVYIMTPYLILDSELEHAIOFAAERGVDVRI

ETILNQTYLSPSLSMTKGDGGVQMIASGPDTRWEVNKKLFFSMITSAKKSIWIASPFYIPDDDDILSALKIAALSGIDVRI

1563 1593 1623 1653 1683 1713 1743 1773
IMPGIDPKPIPYALAKTYYOALTKSGVKITYEYTLGEVHSKIFLSDNTKAVVGTINLDYRSLYHHFECAVYLYKVDAIODI

LVPNRPDKRIVFHASRSYFPELLEAGVKVYEYNRGFMHSKIIVDHEIASIGTSNMDMRSFHLNFEVNAYLYRTSSVTKL

400 410 420 430 440 450 460

1803 1833 1863 1893 1923 1953 1983 2013

YRDYMDTLNKSRLVSLKDINNIPKFQKVGIVTKTIAPLI***K***FIENLILKVN*RI*LYLKSKGCILTKLC***T**TVMP*VD

VSDYVYDLEHSNQINFSLFKNRPFHRLIESTSRLLSPLL
480 490 500

SEQ ID 8954 (GBS277d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 18; MW 51kDa), in Figure 151 (lane 17 & 18; MW 51kDa) and in Figure 182 (lane 12; MW 51kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 15 & 16; MW 76kDa) and in Figure 58 (lane 5; MW 87kDa).

GBS277d-His was purified as shown in Figure 235, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2259-

Example 2006

A DNA sequence (GBSx2116) was identified in *S.agalactiae* <SEQ ID 6205> which encodes the amino acid sequence <SEQ ID 6206>. This protein is predicted to be aspartate-semialdehyde dehydrogenase. Analysis of this protein sequence reveals the following:

```

5      Possible site: 42
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9831> which encodes amino acid sequence <SEQ ID 9832>
was also identified.

15  The protein has homology with the following sequences in the GENPEPT database.

      >GP:AAA26850 GB:J02667 aspartate beta-semialdehyde dehydrogenase (EC
      1.2.1.11) [Streptococcus mutans]
      Identities = 261/357 (73%), Positives = 304/357 (85%), Gaps = 1/357 (0%)

20  Query: 1  MGYTVAIVGATGAVGTQMIRQLEQSNLPQIEQVKLLSSRSAGKILHFKDEAIRVEETKE 60
      MGYTVAIVGATGAVGT+MI+QLEQS LP+++V+LLSSRSAGK+L +KD+ + VE TTK+
      Sbjct: 1  MGYTVAIVGATGAVGTRMIQQLQESTLPVDKVRLLSSRSAGKVLQYKDQDVTVELTTKD 60

25  Query: 61  SFYDVDIALFSAGGSISAKFAPYAVKSGAVVDNTSYFRQNPDPVPLVPEVNAHAMIGHN 120
      SF VDIALFSAGGS+SAKFAPYAVK+GAVVDNTS+FRQNPDPVPLVPEVNA+AM HN
      Sbjct: 61  SPEAVDIALFSAGGSVSASFAPYAVKAGAVVDNTSHFRQNPDPVPLVPEVNAYAMDAHN 120

      Query: 121 GIIACPNCSTIQMMIALEPIRQKWGIERVIVSTYQAVSGSGARAVEETKEQLRQVLNDNL 180
      GIIACPNCSTIQMM+ALEPIRQKWG+ RVIVSTYQAVSG+G A+ ET ++++V+ND +
30  Sbjct: 121 GIIACPNCSTIQMMVALEPIRQKWGLSRVIVSTYQAVSGAGQSAINETVREIKEVVNDGV 180

      Query: 181 SPDQLIATVLPSSDQKHYPPIAFNALPQIDIFTDNDYTYEEMKMTLETKKIMEDATIKVS 240
      P + A + P D+KHYPIAFNAL QID+FTDNDYTYEEMKMT ETKKIME+ + VS
      Sbjct: 181 DPKAVHADIFPSGGDKKHYPIAFNALAQIDVFTDNDYTYEEMKMTNETKKIMEEPELPVS 240

35  Query: 241 ATCVRIIPVLSGHSESIYIETKELASISEIKKAIANFPGAVLQDLPSQIYPQAINAVGHR 300
      A CVR+P+L HSE++YIETK++A I E+K AIA FPGAVL+D QIYPQA NAVG R
      Sbjct: 241 AHCVRVPILFSHSEAVYIETKDVAPIEEVKAIAAIFPGAVLEDDIKHQIYPQAANAVGSR 300

40  Query: 301 ETFVGRIRKDLQENGVMWVSDNLLKGAAWNSVQIAETLHKNGLVKPAKELKFEL 357
      TFVGRIRKDLQ ENG+HMWVSDNLLKGAAWNS+ A LH+ GLV+ ELKFEL
      Sbjct: 301 -TFVGRIRKDLQENGIMWVSDNLLKGAAWNSIITANRLHERGLVRSTSELKFEL 356

```

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2007

A DNA sequence (GBSx2117) was identified in *S.agalactiae* <SEQ ID 6207> which encodes the amino acid sequence <SEQ ID 6208>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 22
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.66    Transmembrane    33 - 49 ( 33 - 49)

      ----- Final Results -----
55  bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

-2260-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 500.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2008

A DNA sequence (GBSx2119) was identified in *S.agalactiae* <SEQ ID 6209> which encodes the amino acid sequence <SEQ ID 6210>. Analysis of this protein sequence reveals the following:

10 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3853(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2009

A DNA sequence (GBSx2120) was identified in *S.agalactiae* <SEQ ID 6211> which encodes the amino acid sequence <SEQ ID 6212>. This protein is predicted to be unnamed protein product (clpP). Analysis of this protein sequence reveals the following:

25 Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3883(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 35 A related GBS nucleic acid sequence <SEQ ID 10061> which encodes amino acid sequence <SEQ ID 10062> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6213> which encodes the amino acid sequence <SEQ ID 6214>. Analysis of this protein sequence reveals the following:

40 Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2682(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 175/196 (89%), Positives = 187/196 (95%)

-2261-

Query: 5 MIPVVIEQTSRGERSDIYSRLLKDRIIMLTGQVEDNMANSIIAQLLFDAQDNTKDIYL 64
 MIPVVIEQTSRGERSDIYSRLLKDRIIMLTG VEDNMANS+IAQLLFDAQDNTKDIYL
 Sbjct: 1 MIPVVIEQTSRGERSDIYSRLLKDRIIMLTGPVEDNMANSVIAQLLFDAQDNTKDIYL 60

5 Query: 65 YVNTPGGSVSAGLAIVDTMNFIKSDVQTIVMGMAASMGTIASSGAKGKRFMLPNAEYMI 124
 YVNTPGGSVSAGLAIVDTMNFIK+DVQTIVMGMAASMGTIASSG KGKRFMLPNAEYMI
 Sbjct: 61 YVNTPGGSVSAGLAIVDTMNFIKADVQTIVMGMAASMGTVIASSGKKGKRFMLPNAEYMI 120

10 Query: 125 HQPMGGTGGGTQQSDMAIAAEHLLKTRHTLEKILADNSGQSIEKVHDDAERDRWMSAQET 184
 HQPMGGTGGGTQQ+DMAIAAEHLLKTRH LEKILA N+G++I+++H DAERD WMSA+ET
 Sbjct: 121 HQPMGGTGGGTQQTDMIAIAAEHLLKTRHLEKILAQNAGKTIKQIHKDAERDYWMSAEET 180

Query: 185 LDYGFIDAIMENNNLQ 200
 L YGFID IMENN L+
 15 Sbjct: 181 LAYGFIDEIMENNELK 196

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2010

20 A DNA sequence (GBSx2121) was identified in *S.agalactiae* <SEQ ID 6215> which encodes the amino acid sequence <SEQ ID 6216>. This protein is predicted to be uracil phosphoribosyltransferase (upp). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -0.43 Transmembrane 127 - 143 (127 - 144)
 INTEGRAL Likelihood = -0.06 Transmembrane 72 - 88 (72 - 89)
 INTEGRAL Likelihood = -0.06 Transmembrane 154 - 170 (154 - 170)

----- Final Results -----
 30 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10063> which encodes amino acid sequence <SEQ ID 10064> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase
 [Streptococcus salivarius]
 Identities = 192/209 (91%), Positives = 202/209 (95%)

40 Query: 1 MGKFQVISHPLIQHKLILRRTTTSTKDFRELVDIAMLNGYEVSRDLPLEDVEIQTPVA 60
 MGKFQVISHPLIQHKLILRR TSTKDFRELVDIAMLNGYEVSRDLPLE+VEIQTP+
 Sbjct: 1 MGKFQVISHPLIQHKLILRREDTSTKDFRELVDIAMLNGYEVSRDLPLEEVEIQTPIT 60

45 Query: 61 TTVQKQLAGKKLAIVPILRAGIGMVDGFLSLVPAKVGHIGMYRDEETFQPV EYLVKLPE 120
 TTVQKQL+GKKLAIVPILRAGIGMVDGFLSLVPAKVGHIGMYRDEET +PVEYLVKLPE
 Sbjct: 61 KTVQKQLSGKKLAIVPILRAGIGMVDGFLSLVPAKVGHIGMYRDEETLEPVEYLVKLPE 120

50 Query: 121 DIDQRQIFVVDPLATGGSAILAVDSLKKRGAASIKFVCLVAAPEGVAALQEAHPDVIDY 180
 DIDQRQIFVVDPLATGGSAILAVDSLKKRGAA+IKFVCLVAAPEGV LQ+AHPD+DIY
 Sbjct: 121 DIDQRQIFVVDPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQDAHPDIDY 180

Query: 181 TAALDEKLNEHGYIVPGLGDAGDRLFGTK 209
 TA+LDEKLNE+GYIVPGLGDAGDRLFGTK
 55 Sbjct: 181 TASLDEKLNEHGYIVPGLGDAGDRLFGTK 209

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6217> which encodes the amino acid sequence <SEQ ID 6218>. Analysis of this protein sequence reveals the following:

-2262-

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 72 - 88 (72 - 89)
 INTEGRAL Likelihood = -0.22 Transmembrane 127 - 143 (127 - 144)

----- Final Results -----

bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to uracil phosphoribosyltransferase from *S.salivarius*:

>GP:AAA26890 GB:L07793 uracil phosphoribosyltransferase [Streptococcus salivarius]
 Identities = 191/209 (91%), Positives = 205/209 (97%)

Query: 1 MGKQVISHPLIQHKLSILRRQTTSTKDFRELVNEIAMLMGYEVSRLPLEDVDIQTTPVS 60
 MGK QVISHPLIQHKLSILRR+ TSTKDFRELVNEIAMLMGYEVSRLPLE+V+IQTTP++
 Sbjct: 1 MGKFQVISHPLIQHKLSILRREDTSTKDFRELVNEIAMLMGYEVSRLPLEEVEIQTTPIT 60

Query: 61 KTVQKQLAGKKLAIVPILRAGIGMVDGLLSLVPAAKVGHIGMYRNEETLEPVEYLVKLPE 120
 KTVQKQL+GKKLAIVPILRAGIGMVDG LSLVPAAKVGHIGMYR+EETLEPVEYLVKLPE
 Sbjct: 61 KTVQKQLSGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETLEPVEYLVKLPE 120

Query: 121 DINQRQIFLVDPMPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQEAHPDIDIF 180
 DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQ+AHPDIDI+
 Sbjct: 121 DIDQRQIFVVDPMPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQDAHPDIDIY 180

Query: 181 TAALDDHLNEHGYIVPGLGDAGDRLFGTK 209
 TA+LD+ LNE+GYIVPGLGDAGDRLFGTK
 Sbjct: 181 TASLDEKLNENGYIVPGLGDAGDRLFGTK 209

An alignment of the GAS and GBS proteins is shown below.

Identities = 190/209 (90%), Positives = 201/209 (95%)

Query: 1 MGKFQVISHPLIQHKLSILRRTTTTSTKDFRELVDEIAMLMGYEVSRLPLEDVEIQTTPVA 60
 MGK QVISHPLIQHKLSILRR TTSTKDFREL+EIAMLGYEVSRLPLEDV+IQTTPV+
 Sbjct: 1 MGKQVISHPLIQHKLSILRRQTTSTKDFRELVNEIAMLMGYEVSRLPLEDVDIQTTPVS 60

Query: 61 TTVQKQLAGKKLAIVPILRAGIGMVDGFLSLVPAAKVGHIGMYRDEETFQPVVEYLVKLPE 120
 TVQKQLAGKKLAIVPILRAGIGMVDG LSLVPAAKVGHIGMYR+EET +PVEYLVKLPE
 Sbjct: 61 KTVQKQLAGKKLAIVPILRAGIGMVDGLLSLVPAAKVGHIGMYRNEETLEPVEYLVKLPE 120

Query: 121 DIDQRQIFVVDPMPLATGGSAILAVDSLKKRGAASIKFVCLVAAPEGVAALQEAHPDVIDIY 180
 DI+QRQIF+VDPMLATGGSAILAVDSLKKRGAA+IKFVCLVAAPEGV LQEAHPD+DI+
 Sbjct: 121 DINQRQIFLVDPMPLATGGSAILAVDSLKKRGAANIKFVCLVAAPEGVKKLQEAHPDIDIF 180

Query: 181 TAALDEKLNNEHGYIVPGLGDAGDRLFGTK 209
 TAALD+ LNEHGYIVPGLGDAGDRLFGTK
 Sbjct: 181 TAALDDHLNEHGYIVPGLGDAGDRLFGTK 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2011

A DNA sequence (GBSx2122) was identified in *S.agalactiae* <SEQ ID 6219> which encodes the amino acid sequence <SEQ ID 6220>. This protein is predicted to be hemolysin (patB). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.29 Transmembrane 88 - 104 (86 - 106)

----- Final Results -----

-2263-

bacterial membrane --- Certainty=0.2317(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15133 GB:Z99120 aminotransferase [Bacillus subtilis]

Identities = 130/381 (34%), Positives = 221/381 (57%), Gaps = 4/381 (1%)

10 Query: 5 DFTSLPERFSSNTIKWKAVQK---DQEILPLWIADMFPIFPEMSEAIEDFSHQMVFGYD 61
 +F ER + ++KW + + LP+W+AD MDF ++EA+++ +FGY
 Sbjct: 2 NFDKREERLGTQSVKWDKTGELFGVTDALPMWVADMDFRAPAEITEALKERLDHGIFGYT 61

15 Query: 62 SPKDSLYQAISNWEVQEHGYQFDKKSLLLIDGVVPAISVAIQAFTEKGDVAVLINTPVYPP 121
 +P A+ W HG++ + +S+ GVV A+S+A+QAFT+ GD V++ PVY P
 Sbjct: 62 TPDQKTKDAVCGWMQNRHGKVNPEISITFSPGVVTALSMVQAFTTEPGDQVVVQPPVYTP 121

20 Query: 122 FARTIKYNNRHLVSNLSLNNNQYFEIDFKQLEKDIIENNVKLYIFCSPHNPGRVWTKGE 181
 F ++ N RH++ N LL + + IDF+ LE + + +V L+I C+PHNP GR W++ +
 Sbjct: 122 FYHMEKNGRHILHNPLEKDGAYIDFEDLETKLSDPSVTLFILCNPHNPSGRSWSRED 181

25 Query: 182 IQKIGDICKRYNVILVSDIEHQDLVLFNDNVHHSFNTVDSSFKELSVILSSATKTFNIAGT 241
 + K+G++C + V +VSDEIH DL+L+ + H F ++ F ++SV ++ +KTFNIAG
 Sbjct: 182 LLKLGEELCEHGVTVVSDIEHSDLMLYGHKHTPFASLSDDFADISVTCAPS KTFNIAGL 241

30 Query: 242 KNSFAIIEENEKLRSDFKKRQIANNOQEISSLGLLATEVAFTKEKQWLKALKMELEGSIEY 301
 + S II + R+ F N +++ + A E A++K WL L +E ++
 Sbjct: 242 QASAIIPDRLKRAKFSASLQRNGLGGLNAFAVTAIEAAYS KGGPWLELITYIEKNMNE 301

35 Query: 302 LYEQL-TQKTHIKVMKPEGTYLWVLDIFSAYNLTHLEIQEKLRYDAKLILNDGLTFGKEGK 360
 L T+ +K+MKP+ +YL+WLDIFSAY L+ E+Q+++ K+IL G +G G+
 Sbjct: 302 AEAFSLSTELPKVKMMKPDASYLIWLDIFSAYGLSDAELQQRMLKKGKVVILEPGTKYGPGE 361

Query: 361 KHARINVAAPRSVIEEAVLRL 381
 R+N + +++ + R+
 Sbjct: 362 GFMRLNAGCSLATLQDGLRRI 382

There is also homology to SEQ ID 1006.

SEQ ID 6220 (GBS392) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 2; MW 46.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 5; MW 71kDa).

GBS392-GST was purified as shown in Figure 217, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2012

45 A DNA sequence (GBSx2123) was identified in *S.agalactiae* <SEQ ID 6221> which encodes the amino acid sequence <SEQ ID 6222>. This protein is predicted to be rRNA methylase, SpoU family (cspR). Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1436(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-2264-

>GP:AAB02738 GB:U58864 CspR [Bacillus subtilis]

Identities = 84/155 (54%), Positives = 120/155 (77%), Gaps = 3/155 (1%)

Query: 19 HIVLFEPQIPANTGNIARTCAATNAPLHIIRPMGFPIDDKKMKRAGLDYWDKLDVSFYDG 78
 H+VL++P+IPANTGNIARTCAATN LH+IRP+GF DDK +KRAGLDYW+ ++V ++D
 Sbjct: 4 HVVLYQPEIPANTGNIARTCAATNTTLHLIRPLGFSTDDKMLKRAGLDYWEFVNNVYHDS 63

Query: 79 LEE-FMLSCRGKVHLISKFADKVYSDENYND-DQDHYFMFGREDKGLPETFMREHAEKAL 136
 LEE F +GK I+KF + ++ +Y D D+D++F+FGRE GLP+ ++ + ++ L
 Sbjct: 64 LEELFEAYKKGKFFFTTKFGQQPHTSFYDITLDEYFFVFGRETSGLPKDLIQNNMDRCL 123

Query: 137 RIPMNDHVRSNLVSNVTVCMIVYEALRQOSFPNLE 171
 R+PM EHVRSLN+SNT ++VYEALRQ++ +L+
 Sbjct: 124 RLPMT-EHVRSLNLSNTAAILVYEALRQQNYRDLK 157

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6223> which encodes the amino acid sequence <SEQ ID 6224>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2236(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 135/182 (74%), Positives = 150/182 (82%)

Query: 1 MNIETLTQKNHRSDSGRNHIVLFEPQIPANTGNIARTCAATNAPLHIIRPMGFPIDDKKM 60
 M + L KN + RNHIVLF+PQIP NTGNIARTCAATNAPLHII+PMGFPIDD+KM
 Sbjct: 13 MTTKELINKNDKVKKARNHIVLFQIPQNTGNIARTCAATNAPLHIKPMGFPIDDRKM 72

Query: 61 KRAGLDYWDKLDVSFYDGLLEEFMLSCRGKVHLISKFADKVYSDENYNDQDHYFMFGRED 120
 KRAGLDYWDKL++ FYD LE+F+ C G++HLISKFA YS Y D HYF+FGRED
 Sbjct: 73 KRAGLDYWDKLELHFYDHLQFINQCHGQLHLISKFAVNNYSQATYADGDSHYFLFGRED 132

Query: 121 KGLPETFMREHAEKALRIPMNDHVRSNLVSNVTVCMIVYEALRQOSFPNLELSHTYENDK 180
 GLPE FMREHAEKALRIPMNDHVRSNLVSNVTVCM++YEALRQ F LEL HTYE+DK
 Sbjct: 133 TGLPEDFMREHAEKALRIPMNDHVRSNLVSNVTVCMVIYEALRQQFGQLELKHITYEHDK 192

Query: 181 LK 182
 LK
 Sbjct: 193 LK 194

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2013

A DNA sequence (GBSx2124) was identified in *S.agalactiae* <SEQ ID 6225> which encodes the amino acid sequence <SEQ ID 6226>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.79	Transmembrane	82 - 98 (69 - 100)
----------	--------------------	---------------	---------------------

INTEGRAL	Likelihood = -6.48	Transmembrane	27 - 43 (24 - 47)
----------	--------------------	---------------	--------------------

INTEGRAL	Likelihood = -5.52	Transmembrane	132 - 148 (126 - 151)
----------	--------------------	---------------	------------------------

INTEGRAL	Likelihood = -5.10	Transmembrane	162 - 178 (161 - 185)
----------	--------------------	---------------	------------------------

----- Final Results -----

bacterial membrane --- Certainty=0.3718(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2265-

A related GBS nucleic acid sequence <SEQ ID 9411> which encodes amino acid sequence <SEQ ID 9412> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:CAB13143 GB:Z99110 similar to amino acid permease [Bacillus subtilis]
    Identities = 46/143 (32%), Positives = 81/143 (56%), Gaps = 1/143 (0%)

    Query: 3  FAYDGTIFVNIAPKPNPKKNLPLAFVIGPALILLSYLAFFYGLTQILGASFIMTTGND 62
              FAYDGW + + E+KNP+K LP A G ++ Y+ + L IL A+ I+T G +
10  Sbjct: 203 FAYDGIWILLALGGMKNPEKLLPRAMTGGLLIVTAIYIFINFALLHILSANEIVTLGEN 262

    Query: 63 AINYAANIIFGPSVGRLLSFIVILSVLGVANGLLGLTMRPQAFERGWIK-SERMANIN 121
              A + AA ++FG G+L+S +I+S+ G NG +L R+ A AER + +E+++++
15  Sbjct: 263 ATSTAATMLFGSIGGKLISVGIIVSIFGCLNGKVLSPFRVSFAMAERKQLPFAEKLSHVH 322

    Query: 122 LKYQMSLPASLTVTAVAIFWLFV 144
              ++ A A+A+ + +
20  Sbjct: 323 PSFRTPWIAISFQIALALIMMLI 345

```

20 There is also homology to SEQ ID 3114.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2014

25 A DNA sequence (GBSx2125) was identified in *S.agalactiae* <SEQ ID 6227> which encodes the amino acid sequence <SEQ ID 6228>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

```

```

30  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1849(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

35 A related GBS nucleic acid sequence <SEQ ID 9439> which encodes amino acid sequence <SEQ ID 9440> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:AAD23454 GB:AF117741 cochaperonin GroES [Streptococcus pneumoniae]
    Identities = 31/52 (59%), Positives = 42/52 (80%)

    Query: 2  GDGIRTLTGELVAPSVAECDTVLVENGAGLEVKGNEKVTVVRESDIVAVVK 53
              G G+RTL G+LVAPSV GD VLVE AGL+VKDG+EK +V E++I+A+++
    Sbjct: 42 GQGVRTLNGDLVAPSVKTDGDRVLVEAHAGLDVKGDEKYIIVGEANILAIIE 93

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6229> which encodes the amino acid sequence <SEQ ID 6230>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have no N-terminal signal sequence

```

```

50  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3290(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-2266-

An alignment of the GAS and GBS proteins is shown below.

Identities = 29/49 (59%), Positives = 39/49 (79%)

```

Query: 4  GIRTLTGELVAPSVAEGLTVLVENGAGLEVKGNEKVTVVRESDIVAVV 52
          G+RT+TG+ V PSV+ G VLVENG LEV +EKV+++RESDI+A+V
Sbjct: 60  GVRTITGDSVLPSVSVGQEVLEVNGHDLVTVDEKVSIIRESDIIAIV 108

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2015

A DNA sequence (GBSx2126) was identified in *S.agalactiae* <SEQ ID 6231> which encodes the amino acid sequence <SEQ ID 6232>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1272 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAD23455 GB:AF117741 chaperonin GroEL [Streptococcus pneumoniae]
Identities = 472/539 (87%), Positives = 513/539 (94%), Gaps = 1/539 (0%)

```

```

Query: 1  MAKDIKFSADARSAMVRGVDILADTVKVTLGPKGRNVVLEKAFGSPPLITNDGVTIAKEIE 60
          M+K+IKFS+DARSAMVRGVDILADTVKVTLGPK RNVVLEK+FGSPPLITNDGVTIAKEIE
Sbjct: 1  MSKEIKFSSDARSAMVRGVDILADTVKVTLGPKDRNVVLEKSFSGSPPLITNDGVTIAKEIE 60

```

```

Query: 61  LEDHFENMGAKLVSEVASKTNDIAGDGTITATVLTQAIIVREGLKNVTAGANPIGIRRGIE 120
          LEDHFENMGAKLVSE+ASKTNDIAGDGTITATVLTQAIIVREG+KNVTAGANPIGIRRGIE
Sbjct: 61  LEDHFENMGAKLVSEIASKTNDIAGDGTITATVLTQAIIVREGIKNVTAGANPIGIRRGIE 120

```

```

Query: 121 TAVSAAVEELKEIAQPVSGKEAIAQVAAVSSRSEKVGGEYISEAMERVGNMGVITIEESRG 180
          TAV+AAVE LK A PV+ KEAI+QVAAVSSRSEKVGGEYISEAME+VG DGVITIEESRG
Sbjct: 121 TAVAAAVEALKNNAIPVANKEAISQVAAVSSRSEKVGGEYISEAMEKVGKDGVTIEESRG 180

```

```

Query: 181 METELEVVVEGMQFDRGYLSQYMVTDNEKVMSELENPYILITDKKISNIQEILPLEEVLEK 240
          METELEVVVEGMQFDRGYLSQYMVTD+EKMV++LENPYILITDKKISNIQEILPLEE +L+
Sbjct: 181 METELEVVVEGMQFDRGYLSQYMVTDSEKVMADLENPYILITDKKISNIQEILPLESILQ 240

```

```

Query: 241 TNRPLLI IADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGTVVT 300
          +NRPLLI IADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGTV+T
Sbjct: 241 SNRPLLI IADDVDGEALPTLVLNKIRGTFNVVAVKAPGFGDRRKAMLEDIAILTGTVIT 300

```

```

Query: 301 EDLGLDLKDATMQVLGQSAKVTVDKDSTVIVEGAGDSSAIANRVAIIKSQMEATTSDFDR 360
          EDLGL+LKDAT++ LGQ+A+VTVDKSTVIVEGAG+ AI++RVA+IKSQ+E TTS+FDR
Sbjct: 301 EDLGLDLKDATIEALGQAARVTVDKDSTVIVEGAGNPEAISHRVAVIKSQIETTTSEFDR 360

```

```

Query: 361 EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVSGGGTALVNV 420
          EKLQERLAKL+GGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIV+GGGTAL NV
Sbjct: 361 EKLQERLAKLSGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVAGGGTALANV 420

```

```

Query: 421 IEKVAALKINGDEETGRNIVLRALEEPVRQIAYNAGYEGSVIIRLKQSEIGTGFNAAANG 480
          I A L+L GDE TGRNIVLRALEEPVRQIA+NAG+EGS++I+RLK +E+G GFNAA G
Sbjct: 421 IPAEATLELTGDEATGRNIVLRALEEPVRQIAHNAGFEFSIVIDRLKNAELGIGFNAAATG 480

```

```

Query: 481 EWVDMVTTGIIDPVKVTRESALQNAASVASLILTEAVVANKPEPEAPTAPAMDPSMMGG 539
          EWV+M+ GIIDPVKV+RSALQNAASVASLILTEAVVANKPEP AP APAMDPSMMGG
Sbjct: 481 EWNMIDQGIIDPVKVSRSALQNAASVASLILTEAVVANKPEPVAP-APAMDPSMMGG 538

```

-2267-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6233> which encodes the amino acid sequence <SEQ ID 6234>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1070(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 491/543 (90%), Positives = 515/543 (94%), Gaps = 3/543 (0%)

```

Query: 1  MAKDIKFSADARSAMVRGVDILADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 60
          MAKDIKFSADAR+AMVRGVD+LADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE
Sbjct: 3  MAKDIKFSADARAAMVRGVDMLADTVKVTLGPKGRNVVLEKAFGSPLITNDGVTIAKEIE 62

Query: 61  LEDHFENMGAKLVSEVASKTNDIAGDGTITATVLTQAIVREGLKNVTAGANPIGIRRGIE 120
          LEDHFENMGAKLVSEVASKTNDIAGDGTITATVLTQAIV  EGLKNVTAGANPIGIRRGIE
Sbjct: 63  LEDHFENMGAKLVSEVASKTNDIAGDGTITATVLTQAIVHEGLKNVTAGANPIGIRRGIE 122

Query: 121  TAVSAAVEELKEIAQPVSGKEAIAQVAAVSSRSEKVG EYISEAMERVGN DGVITIEESRG 180
          TA + AVE LK  IAQPVSGKEAIAQVAAVSSRSEKVG EYISEAMERVGN DGVITIEESRG
Sbjct: 123  TATATAVEALKIAIAQPVSGKEAIAQVAAVSSRSEKVG EYISEAMERVGN DGVITIEESRG 182

Query: 181  METELEVVVEGMQFDRGYLSQYMTDNEKMVSELENPFILITDKKISNIQEILPLLEEVLK 240
          METELEVVVEGMQFDRGYLSQYMTDNEKMV++LENP+ILITDKK+SNIQ+ILPLLEEVLK
Sbjct: 183  METELEVVVEGMQFDRGYLSQYMTDNEKMVADLENPFILITDKKVSNIQDILPLLEEVLK 242

Query: 241  TNRPLLIIADDVDGEALPTLVLNKIRGTFN VAVKAPGFGDRRKAMLEDIAILTG GTVVT 300
          TNRPLLIIADDVDGEALPTLVLNKIRGTFN VAVKAPGFGDRRKAMLEDIAILTG GTV+T
Sbjct: 243  TNRPLLIIADDVDGEALPTLVLNKIRGTFN VAVKAPGFGDRRKAMLEDIAILTG GTVIT 302

Query: 301  EDLGLDLKDATMQVLGQSAKVTVDKDSTVIVEGAGDSSAIA NRVAIIKSQMEATTSD FDR 360
          EDLGL+LKDATM  LGQ+AK+TVDKDSTVIVEG+G S AIA NR+A+IKSQ+E TTSDFDR
Sbjct: 303  EDLGLBLKDATMTALGQA AKITVDKDSTVIVEGSGSSEAIA NRIALIKSQLETTTSD FDR 362

Query: 361  EKLQERLAKLAGGVAVIKVGAATETELKEMKLRIEDALNATRAAVEEGIVSGGGTALVNV 420
          EKLQERLAKLAGGVAVIKVGA TET LKEMKLRIEDALNATRAAVEEGIV+GGGTAL+ V
Sbjct: 363  EKLQERLAKLAGGVAVIKVGAPTETALKEMKLRIEDALNATRAAVEEGIVAGGGTALITV 422

Query: 421  IEKVAALKLNGDEETGRNIVLRALEEPVRQIAYNAGYEGSVIIERLKQSEIGTGFNAANG 480
          IEKVAAL+L GD+ TGRNIVLRALEEPVRQIA NAGYEGSV+I++LK S GTGFNAAG G
Sbjct: 423  IEKVAALELEGDDATGRNIVLRALEEPVRQIALNAGYEGSVVIDKLKNSPAGTG FNAATG 482

Query: 481  EWVDMVTGTGIIDPVKVTR SALQNAASVASLILTEAVVANKPEP--EAPTAPA-MDPSMM 537
          EWVDM+ TGIIDPVKVTR SALQNAASVASLILTEAVVANKPEP  AP PA MDP MM
Sbjct: 483  EWVDMIKTGTGIIDPVKVTR SALQNAASVASLILTEAVVANKPEPATPAPAMPAGMDPGMM 542

Query: 538  GGF 540
          GGF
Sbjct: 543  GGF 545

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2016

A DNA sequence (GBSx2127) was identified in *S.agalactiae* <SEQ ID 6235> which encodes the amino acid sequence <SEQ ID 6236>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

-2268-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3216(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10247> which encodes amino acid sequence <SEQ ID 10248> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06113 GB:AP001515 transcriptional regulator (GntR family)
 [Bacillus halodurans]
 Identities = 50/171 (29%), Positives = 86/171 (50%), Gaps = 17/171 (9%)

Query: 21 HVQVYNKIFNMIQDGTYSFGMQLPSEPELAGQLNVSRAITLRKSLALLQEDHLVKNIRGKG 80
 ++QV +K+ + ++ G Y G +LPSE EL+ QL VSRATLR++L LL+E+ +V G G
 Sbjct: 10 YLQVIDKLKHDMEAGVYEEGEKLPSEFELSKQLGVSRAITLRREALRLLEEGLVVRHGVG 69

Query: 81 NFIRENSSNLSENGYENRQHPIKTCLTSKITEVELE-----FRVEVPAEAITASLKQ 132
 F+ ++ L G E +T I ++E +++E +
 Sbjct: 70 TFV--HTKPLFSAGIEELY-----SVTDMIRHADMEPGTIFLSSYQIEATDDDKRRFQTD 122

Query: 133 ETPVVVIADRWYHTDDGPLAYTSLFPIELISDAEISLHDTKQLNFIIEG 183
 +++ +R D P+ Y L +P ELI + S+H+ +L+ +E G
 Sbjct: 123 NLDQLMMIERVVRTADGVPIVYCLDKLPAELI--GQHSVHEINSILDHLESG 171

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6237> which encodes the amino acid sequence <SEQ ID 6238>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2297(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 154/244 (63%), Positives = 189/244 (77%)

Query: 7 MPKNELNNKLNKLKHVQVYNKIFNMIQDGTYSFGMQLPSEPELAGQLNVSRAITLRKSLAL 66
 M N+L KL KLKHVQVYN IF +IQDGTYSFGMQLPSEPELA QLNVSRAITLRKSLAL
 Sbjct: 1 MSTNDLTKKLKLKHVQVYNTIFQLIQDGTYSFGMQLPSEPELAGQLNVSRAITLRKSLAL 60

Query: 67 LQEDHLVKNIRGKGNFIRENSSNLSENGYENRQHPIKTCLTSKITEVELEFRVEVPAEAI 126
 LQEDHL+KNIRGKGNFI + G+E QHPI L+S IT+VELE+R+EVP AI
 Sbjct: 61 LQEDHLIKNIRGKGNFILKTPETKYHQGFYQLQHPIYASLSSDITKVELEYRIEVPVTAI 120

Query: 127 TASLKQETPVVVIADRWYHTDDGPLAYTSLFPIELISDAEISLHDTKQLNFIIEGIIYQ 186
 TASLKQETPVV+I DRWYH+ + +AY+LSFIPIE+IS I+L+ + LL F+EE IY+
 Sbjct: 121 TASLKQETPVVVIIDRWYHSQNKAIAYSLSFIPIEVISKYAINLNQEEPLLTFLLEKIYE 180

Query: 187 EGISHSQSHLGYATSGNFSATKYTLSDHGQFILIQETIFKQEKILMCNKHYVPIEHFEL 246
 G +SHS + +GY +GN++ATKYTL++ FILIQUET++ + IL+ KHYVP + F+L
 Sbjct: 181 SGKASHSCNQIGYTKTGNATKYTLSENSAFILIQUETLYNGKDILVSTKHYVPADLFDL 240

Query: 247 SITS 250
 + S
 Sbjct: 241 KVQS 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2269-

Example 2017

A DNA sequence (GBSx2128) was identified in *S.agalactiae* <SEQ ID 6239> which encodes the amino acid sequence <SEQ ID 6240>. This protein is predicted to be purine nucleoside phosphorylase (udp-1). Analysis of this protein sequence reveals the following:

```

5   Possible site: 47
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3910(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAC65977 GB:AE001270 uridine phosphorylase (udp) [Treponema
    pallidum]
    Identities = 145/246 (58%), Positives = 171/246 (68%)

    Query: 11 QYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYTGTLNGEKVSVTST 70
    +YH+ ++ D+G YVI+PGDP R KIA+HF + V +REYVTTYTGTL VSV ST
20   Sbjet: 10 EYHIGLKASDIGHYVILPGDPARSEKIAQHFSHPKHVGHNREYVTTYTGTLCTPVSVMST 69

    Query: 71 GIGGPSASIAM EELKLCGADTFIRVGTCTGGIDLDVKGGDIVIATGAIRMEGTSKEYAPIE 130
    GIGGPS +I +EEL GA TFIRVGT GG+ D+ G +VIATGAIR EGTSKEYAP+E
25   Sbjet: 70 GIGGPSTAIGVEELIHLGAHTFIRVGTSGGMQPDILAGTVVIATGAIRFEGTSKEYAPVE 129

    Query: 131 FPAVADLEVTNALVNAAKKGYSHTAGVQCKDAFYGQHEPERMPVSYELLNKWEAWKRL 190
    FPAV D VT AL +AA+ + GVVQCKD FYGQH P MPV EL KW AW
    Sbjet: 130 FPAVPDFTVTAALKHAAEDVQVRHALGVVQCKDNFYGQHSPTMPVHAELTQKWHAWIAC 189

30   Query: 191 GTKASEMES AALFVAASHLGVRCGSDFLVGNQERNALGMDNPM AHDTEAAIQVAVEALR 250
    T ASEMESAALFV S VR G+ LV+GNQ R A G+++ HDTE AI+VAVEA++
    Sbjet: 190 NTLASEMES AALFVLGSVRRVRTGAVLLVIGNQTRRAQGLEDIQVHDTENAIRVAVEAVK 249

35   Query: 251 TLIEND 256
    LI D
    Sbjet: 250 LLITQD 255

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6241> which encodes the amino acid sequence <SEQ ID 6242>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 47
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45          bacterial cytoplasm --- Certainty=0.3910(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

50   Identities = 259/259 (100%), Positives = 259/259 (100%)

    Query: 1  MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYTGTL 60
    MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYTGTL
    Sbjet: 1  MQNYSGEVGLQYHLQIRPGDVGRYVIMPGDPKRCAKIAEHFDNAVLVADSREYVTTYTGTL 60

55   Query: 61  NGEKVSVTSTGIGGPSASIAM EELKLCGADTFIRVGTCTGGIDLDVKGGDIVIATGAIRME 120
    NGEKVSVTSTGIGGPSASIAM EELKLCGADTFIRVGTCTGGIDLDVKGGDIVIATGAIRME
    Sbjet: 61  NGEKVSVTSTGIGGPSASIAM EELKLCGADTFIRVGTCTGGIDLDVKGGDIVIATGAIRME 120

60   Query: 121 GTSKEYAPIEFPAVADLEVTNALVNAAKKGYSHTAGVQCKDAFYGQHEPERMPVSYEL 180
    GTSKEYAPIEFPAVADLEVTNALVNAAKKGYSHTAGVQCKDAFYGQHEPERMPVSYEL
    Sbjet: 121 GTSKEYAPIEFPAVADLEVTNALVNAAKKGYSHTAGVQCKDAFYGQHEPERMPVSYEL 180

```

-2270-

Query: 181 LNKWEAWKRLGTKASEMESAAALFVAASHLGVRGSDFLVVGNGQERNALGMDNPMAMDTEA 240
 LNKWEAWKRLGTKASEMESAAALFVAASHLGVRGSDFLVVGNGQERNALGMDNPMAMDTEA
 Sbjct: 181 LNKWEAWKRLGTKASEMESAAALFVAASHLGVRGSDFLVVGNGQERNALGMDNPMAMDTEA 240

Query: 241 AIQVAVEALRTLIENTDKSQ 259
 AIQVAVEALRTLIENTDKSQ
 Sbjct: 241 AIQVAVEALRTLIENTDKSQ 259

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2018

A DNA sequence (GBSx2129) was identified in *S.agalactiae* <SEQ ID 6243> which encodes the amino acid sequence <SEQ ID 6244>. This protein is predicted to be nucleoside transporter. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	35 - 51 (30 - 57)
INTEGRAL	Likelihood = -9.29	Transmembrane	8 - 24 (1 - 28)
INTEGRAL	Likelihood = -8.07	Transmembrane	388 - 404 (379 - 404)
INTEGRAL	Likelihood = -7.27	Transmembrane	104 - 120 (100 - 127)
INTEGRAL	Likelihood = -6.58	Transmembrane	259 - 275 (255 - 284)
INTEGRAL	Likelihood = -4.35	Transmembrane	172 - 188 (171 - 190)
INTEGRAL	Likelihood = -3.50	Transmembrane	200 - 216 (199 - 221)
INTEGRAL	Likelihood = -2.18	Transmembrane	352 - 368 (352 - 371)

----- Final Results -----
 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10245> which encodes amino acid sequence <SEQ ID 10246> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
 Identities = 160/405 (39%), Positives = 256/405 (62%), Gaps = 8/405 (1%)

Query: 5 MQFIYSIIGILLVLGIVYAI SFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQVSVVS 64
 M ++ ++GI++V I +A S NR+++ I L +Q + A+I+++IP GQ ++ ++
 Sbjct: 1 MNILWGLLGIVVFLIAFAFSTNRRRAIKPRTILGGLAIQLLFAIIVLKIPAGQALLESILT 60

Query: 65 TGVTKVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLLYVVGILGFVV 124
 V +I+ G++FVFG + G+ GF+FAI L ++F SAL+S+LYY+GI+ FV+
 Sbjct: 61 NVVLNIISYANEGIDFVFGGFEEGSGVGFVFAINVLSVVIFFSALISTILYYLGMQFVI 120

Query: 125 KWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGSMS 184
 K IG + ++ +S+ ES A AN+F+GQT++P++V YL +MT SE+ V+ G+ S++
 Sbjct: 121 KIIGGALSWLLGTSKAESMSAAANIFVQTEAPLVVKPYLPKMTQSELFVMTGGLASVA 180

Query: 185 VSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKI-DDIKMDNKGNNANV 243
 S+L GY LG+P++YLL AS M +++AK+++P+TE DD K+ + N+
 Sbjct: 181 GSVLIGYSLLGVPLQYLLAASFMAAPAGLIMAKMIMPETEKTDAEDDFKLAKDEESTNL 240

Query: 244 IDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-----IRLEQIFS YVFAP 296
 IDA A GASTG + +I A L+AFV L++LIN +L +G + LE I YVFAP
 Sbjct: 241 IDAAANGASTGLMLVLNIAAMLLAFVALIALINGILGWIGGLFGASQLSLELILGYVFAP 300

Query: 297 FGFLMGFDHKNILLEGNLLGSKLILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLS 356
 F++G L G+ +G KL++NEFV++ I++L + +V + +LCGFAN S

-2271-

Sbjct: 301 LAFVIGIPWAEALQAGSYIGQKLVVNEFVAYLSFAPEIENLSDKAVMVISFALCGFANFS 360

Query: 357 SLGICVSGIAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIV 401
 SLGI + G+ L P +R +ARL RA++ G S+LSA I G++

Sbjct: 361 SLGILLGGLGKLAPSRRPDIARLGLRAILAGTLASLLSASIAGML 405

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6245> which encodes the amino acid sequence <SEQ ID 6246>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	35 - 51 (30 - 57)
INTEGRAL	Likelihood = -9.29	Transmembrane	8 - 24 (1 - 28)
INTEGRAL	Likelihood = -8.07	Transmembrane	388 - 404 (379 - 404)
INTEGRAL	Likelihood = -7.27	Transmembrane	104 - 120 (100 - 127)
INTEGRAL	Likelihood = -6.58	Transmembrane	259 - 275 (255 - 284)
INTEGRAL	Likelihood = -4.35	Transmembrane	172 - 188 (171 - 190)
INTEGRAL	Likelihood = -3.50	Transmembrane	200 - 216 (199 - 221)
INTEGRAL	Likelihood = -2.18	Transmembrane	352 - 368 (352 - 371)

----- Final Results -----

bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB05165 GB:AP001512 nucleoside transporter [Bacillus halodurans]
 Identities = 160/405 (39%), Positives = 257/405 (62%), Gaps = 8/405 (1%)

Query: 5 MQFIYSIIIGILLVLGIVYAI SFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQIVSVVS 64
 M ++ ++GI++V I +A S NR+++ I L +Q + A+I+++IP GQ ++ ++
 Sbjct: 1 MNILWGLLGIVVFLIAFAFSTNRRAIKPRTILGGLAIQLLFAIIVLKIPAGQALLESIT 60

Query: 65 TGVTSVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLYYVGILGFVV 124
 V ++I+ G++FVFG + G+ GF+FAI L ++F SAL+S+LYY+GI+ FV+
 Sbjct: 61 NVVLNIISYANEGIDFVFGGFFEEGSGVGFVFAINVLSVVIFFSALISILYYLGIMQFVI 120

Query: 125 KWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGMSMS 184
 K IG + ++ +S+ ES A AN+F+GQT++P++V YL +MT SE+ V+ G+ S++
 Sbjct: 121 KIIGGALSWLLGTSKAESMSAAANIFVGQTEAPLVVKPYLPKMTQSELFVMTGGLASVA 180

Query: 185 VSILGGYIALGIPMEYLLIASTMVPIGSILIAKILLPQTEPVQKI-DDIKMDNKGNNANV 243
 S+L GY LG+P++YLL AS M +++AK+++P+TE DD K+ + N+
 Sbjct: 181 GSVLIGYSLLGVPLQYLLAASFMAAPAGLIMAKMIMPETEKT'TDAEDDFKLAKDEESTNL 240

Query: 244 IDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-----IRLEQIFSYVFAP 296
 IDA A GASTG + +I A L+AFV L++LIN +L +G + LE I YVFAP
 Sbjct: 241 IDAAANGASTGLMLVLNIAAMLLAFVALIALINGILGWIGGLFGASQLSLELILGYVFAP 300

Query: 297 FGFLMGFDHKNILLEGNLLGSKLIINEFVSFQQLGHLLIKSLDYRTALVATISLCGFANLS 356
 F++G L G+ +G KL++NEFV++ I++L + +V + +LCGFAN S
 Sbjct: 301 LAFVIGIPWAEALQAGSYIGQKLVVNEFVAYLSFAPEIENLSDKAVMVISFALCGFANFS 360

Query: 357 SLGICVSGIAVLCPEKRSTLARLVFRAMIGGIAVSMLSAFIVGIV 401
 SLGI + G+ L P +R +ARL RA++ G S+LSA I G++

Sbjct: 361 SLGILLGGLGKLAPSRRPDIARLGLRAILAGTLASLLSASIAGML 405

An alignment of the GAS and GBS proteins is shown below.

Identities = 399/404 (98%), Positives = 401/404 (98%)

Query: 1 MEVIMQFIYSIIIGILLVLGIVYAI SFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQVV 60
 +EVIMQFIYSIIIGILLVLGIVYAI SFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQ+V
 Sbjct: 1 LEVIMQFIYSIIIGILLVLGIVYAI SFNRKSVSLSLIGKALIVQFIIALILVRIPLGQQIV 60

Query: 61 SVVSTGVTKVINCGQAGLNFVFGSLADSGAKTGFIFAIQTLGNIVFLSALVSLYYVGIL 120

		SVVSTGVT VINCGQAGLNFVFGSLADSGAKTGFI FAIQTGLGNIVFLSALVSLLYYVGIL	
Sbjct:	61	SVVSTGVTSVINCGQAGLNFVFGSLADSGAKTGFI FAIQTGLGNIVFLSALVSLLYYVGIL	120
Query:	121	GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM	180
		GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM	
Sbjct:	121	GFVVKWIGKGVGKIMKSSEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGM	180
Query:	181	GSMSVSILGGYIALGIPMEYLLIASTMPVIGSILIAKILLPQTEPVQKIDDIKMDNKGNN	240
		GSMSVSILGGYIALGIPMEYLLIASTMPVIGSILIAKILLPQTEPVQKIDDIKMDNKGNN	
Sbjct:	181	GSMSVSILGGYIALGIPMEYLLIASTMPVIGSILIAKILLPQTEPVQKIDDIKMDNKGNN	240
Query:	241	ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLGIRLEQIFS YVFAPFGFL	300
		ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLGIRLEQIFS YVFAPFGFL	
Sbjct:	241	ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLGIRLEQIFS YVFAPFGFL	300
Query:	301	MGFDHKNILLEGNNLGSKLIILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLSSLGI	360
		MGFDHKNILLEGNNLGSKLIILNEFVSFQQLGDLIKSLDYRTALVATISLCGFANLSSLGI	
Sbjct:	301	MGFDHKNILLEGNNLGSKLIILNEFVSFQQLGHLIKSLDYRTALVATISLCGFANLSSLGI	360
Query:	361	CVSGIAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIVTLF	404
		CVSGIAVLCPEKR TLARLVFRAMIGGIAVSMLSAFIVGIVTLF	
Sbjct:	361	CVSGIAVLCPEKRSTLARLVFRAMIGGIAVSMLSAFIVGIVTLF	404

Lipop: Possible site: -1 Crend: 1
McG: Discrim Score: 13.83
GvH: Signal Score (-7.5): -2.63
 Possible site: 25

*** Reasoning Step: 3

ORF01622(313 - 1512 of 1812)
 GP|9656920|gb|AAF95495.1||AE004305(1 - 418 of 418) NupC family protein {Vibrio cholerae}
 %Match = 24.0
 %Identity = 39.5 %Similarity = 65.7
 Matches = 160 Mismatches = 134 Conservative Sub.s = 106

VSTGVTKVINCGQAGLNFVFGSLADSG-----AKTGFIFAIQTGLNIVFLSALVSLLYYVGILGFVVKWIGKGVGKIMKS

| | : ||| | | : : ||| | |||| : | :::||| : |||| : :: : ||:: : | : : : :

-2273-

```

FSDAVSNVINYGNDGTSFLFGGLVSGKMFVEVFGGGGFIFAFRVLPTLIFFSALISVLYYLGVWQWVIRILGGGLQKALGT
      70      80      90     100     110     120     130

5   741      771      801      831      861      891      921      951
SEVESFVAVANMFLGQTDSPILVSKYLGRMTDSEIMVVLVSGMGSMSVILGGYIALGIPMEYLLIASTMVPIGSILIAK
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
SRAESMSAAANIFVGQTEAPLVVRPFVVKMTQSELFVAVMCGGLASIAGGVLAGYASMGVKIEYLVAA SFMAAPGGLLFAK
      150     160     170     180     190     200     210

10  981      1011     1038     1068     1098     1128              1167
ILLPQTEPVQKIDDIKMDNKGNN-ANVIDAIAEGASTGAQMAFSIGASLIAFVGLVSLINMMLSGLG-----IRLEQI
:::|:|  |  |:|:|  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
LMPPETEKPQDNEDITLDGGDDKPANVIDAAAGGASAGLQALNVGAMLI AFIGLIALINGMLGGIGGWFGMPPELKLEML
      230     240     250     260     270     280     290

15  1197     1227     1257     1287     1305     1332     1362     1392
FSYVFAPFGFLMGFDHKNILLEGNLLGSKLIILNEFVSFQQ---LGDLIKS-LDYRTALVATISLCGFANLSSLGICVSG
:  ::|:|  |  |:|  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
LGWLFAPLAFLIGVWPNEATVAGEFIGLKTIVANEVAYSQFAPYLTEAAPVVLSEKTKAIISFALCGFANLSSIAILLGG
      310     320     330     340     350     360     370

20  1422     1452     1482     1512     1542     1572     1602     1632
IAVLCPEKRGTLARLVFRAMIGGIAVSMLSAFIVGIVTLF*KLTKERRIVTWK*KIF*KR*TILC*QQQHQGKSKQF*M
:  |  |:|  |  |:|  |  :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
25  LGSLAPKRRGDIARMGVKAVIAGTLSNLMAATIAGFFLSF
      390     400     410

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 2019

A DNA sequence (GBSx2130) was identified in *S.agalactiae* <SEQ ID 6247> which encodes the amino acid sequence <SEQ ID 6248>. This protein is predicted to be deoxyribose-phosphate aldolase (deoC). Analysis of this protein sequence reveals the following:

```

Possible site: 49
35  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2196(Affirmative) < succ>
40      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA81646 GB:Z27121 deoxyribose aldolase [Mycoplasma hominis]
Identities = 99/199 (49%), Positives = 140/199 (69%), Gaps = 1/199 (0%)
45
Query: 5   DILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASVVKAAEYVSGK-LAICTVI 63
++ K +DHT L+ +AT +I ++ +A+ Y+ S CI SYVK A E + + +CTVI
Sbjct: 3   ELNKYIDHTNLSPSATS KDIKLIQEAIKYDFKSVCTIAPSIVKYAKEALKNSDVLVCTVI 62

50  Query: 64  GFPNGYSTTAAKVFECD AIKNGADEIDMVINLTDVKN GDFDTVEEEIRQIKAACQDHIL 123
      GFP GY+ T+ KV+E + A+++GADEIDMVIN+ K+G ++ V EI+ IK AC L
Sbjct: 63  GFPLGYNATSVKVYETKI AVEHGADEIDMVINVG RFDGQY EYVLNEIKAIKEACNGKTL 122

55  Query: 124 KVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMAKYVGEGV KIK A 183
      KVIVET LTK ELI++ +V +SGADFIKTSTGFS GA+FED++ M + G+ + IKA
Sbjct: 123 KVIVETALLTKAELIKITELVMQSGADFIKTSTGFSYRGASFEDIQTMKETCGDKLLIKA 182

Query: 184 AGGISSLEDAEKFIALGAS 202
      +GGI +L DA++ I LGA+
60  Sbjct: 183 SGGIKNLADAQEMIRLGAN 201

```

-2274-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6249> which encodes the amino acid sequence <SEQ ID 6250>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2196(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 211/223 (94%), Positives = 217/223 (96%)

Query: 1 MEVKDILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASYVKKAAEYVSGKLAIC 60
+EVKDILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASYVKKAAEYVSGKLAIC
Sbjct: 1 VEVKDILKTV DHTLLATTATWPEIQ TILDDAMAYETASACIPASYVKKAAEYVSGKLAIC 60

Query: 61 TVIGFPPNGYSTTAAKVFECD AIKNGADEIDMVINLTDVKN GDFDTVEEEIRQIKAACQD 120
TVIGFPPNGYSTTAAKVFECD AIKNGADEIDMVINLTDVKN GDFDTVEEEIRQIK A CQD
Sbjct: 61 TVIGFPPNGYSTTAAKVFECD AIKNGADEIDMVINLTDVKN GDFDTVEEEIRQIKAKCQD 120

Query: 121 HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMKYVGEGVK 180
HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMKYVGEGVK
Sbjct: 121 HILKVIVETCQLTKEELIELCGVVTRSGADFIKTSTGFSTAGATFEDVEVMKYVGEGVK 180

Query: 181 IKAAGGISSLEDAEKFIALGASRLGTSRIIKIVKNQKVEEGTY 223
IKAAGGISSLEDA+ FIALGASRLGTSRIIKIVKN+ + +Y
Sbjct: 181 IKAAGGISSLEDAKTFIALGASRLGTSRIIKIVKNEATKTDSY 223

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2020

A DNA sequence (GBSx2131) was identified in *S.galactiae* <SEQ ID 6251> which encodes the amino acid sequence <SEQ ID 6252>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0546(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45496 GB:U80410 phosphopentomutase [Lactococcus lactis subsp.
cremoris]

Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3 QFDRHILVVLDSVGIGAAPDANDFVNAGVP-----DGASDTLGHISKTVGLAVPNMAKI 56
+F RIHLVV+DSVGIGAAPDA+ F N V D SDT+GHIS+ GL VPN+ K+
Sbjct: 4 KFGRIHLVVMDSVGIGAAPDADKFFNHDVETHEAINDVKSDTIGHISEIRGLDVPNLQKL 63

Query: 57 GLGNIPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWIEMGLNITEPFDTFWNGFP 116
G GNIPR LKT+PA + P+ Y TKL+E+S GKDTMTGHWIEMGLNI PF T+ G+P
Sbjct: 64 GWGNIPRESPLKTIPAAQKPAAYVTKL E EISKGKDTMTGHWIEMGLNIQT PFPPTYPEGY P 123

Query: 117 EDIITKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIIYTSADPVLQIAAHEDI 176
ED++ KIE+FSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELIIYTSADPVLQIAAHED+

-2275-

Sbjct: 124 EDLLEKIEEFSGRKIIREANKPYSGTAVIEDFGPRQLETGELIITYTSADPVLQIAAHEDV 183

Query: 177 IPLEELYRICEYARSITMERPALL-GRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLN 235
I EELY+ICEY RSIT+E ++ GRIIARPYVGE GNF RT R DYA+SPF +TVL

5 Sbjct: 184 ISREELYKICEYVRSITLLEGGSGIMIGRIIARPYVGEAGNFERTDGRRDYALSPFAETVLE 243

Query: 236 KLDQAGIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLV 295
KL +AGIDTY+VGKI+DIEN G+ +DMGHN ++ G+D L+K M +EF +GFSFTNLV

10 Sbjct: 244 KLYKAGIDTYSVGKISDIFNTVGKYMGMHNDMDGVDRLKAMTKTEFTGFSFTNLV 303

Query: 296 DFDALYGHRRDPHGYRDCLEHFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREY 355
DFDA YGHRRD GY + +FD RLPEII AM++ DLL+ITADHGNDP+Y GTDHTREY

Sbjct: 304 DFDKAGHRRDVEGYGKAIEDFDGRLPEIIDAMKEDDLLMITADHGNDPSYVGTHTREY 363

15 Query: 356 IPLLAYSPTSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDLV 403
IPL+ +S SF ++PVGHFADISAT+A+NF V A GESFL LV

Sbjct: 364 IPLVIFSKSFKEPKVLPVGHFADISATIAENFSVKKQTGESFLDALV 411

There is also homology to SEQ ID 2740:

20 Identities = 348/402 (86%), Positives = 374/402 (92%)

Query: 1 MSQFDRIHLVVLDSVGIGAAPDANDFNAGVPDASDTLGHISKTGVLAVPNMAKIGLGN 60
MS+F+RIHLVVLDSVGIGAAPDA+ F NAGV D SDTLGHIS+ GL+VPNMAKIGLGN

25 Sbjct: 1 MSKFNRIHLVVLDSVGIGAAPDADKFFNAGVADTSDTLGHISEAAGLSVPNMAKIGLGN 60

Query: 61 IPRPQALKTVPAEENPSGYATKLEQVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPEDII 120
I RP LKTVP E+NP+GY TKL+EVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPPE+I+

30 Sbjct: 61 ISRPILKTVPTEDNPTGYVTLEEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPPEIL 120

Query: 121 TKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIITYTSADPVLQIAAHEDIIPLE 180
TKIE+FSGRK+IREANKPYSGTAVIDDFGPRQMETGELI+YTSADPVLQIAAHEDIIP+E

35 Sbjct: 121 TKIEEFSGRKIIREANKPYSGTAVIDDFGPRQMETGELIVYTSADPVLQIAAHEDIIPVE 180

Query: 181 ELYRICEYARSITMERPALLGRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLNKLQDA 240
ELY+ICEYARSIT+ERPALLGRIIARPYVG+PGNFTRTANRHDYAVSPF+DTVLNKL A

40 Sbjct: 181 ELYKICEYARSITLERPALLGRIIARPYVGDPGNFTRTANRHDYAVSPFQDTVNLKLADA 240

Query: 241 GIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLVDFDAL 300
G+ TYAVGKINDIFNGSGI +DMGHNKSNSHGIDTLIKT+ L EF KGFSFTNLVDFDA

45 Sbjct: 241 GVPTYAVGKINDIFNGSGITNDMGHNSNSHGIDTLIKTLQLPEFTKGFSFTNLVDFDAN 300

Query: 301 YGHRRDPHGYRDCLEHFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREYIPLLA 360
+GHRRDP GYRDCLEHFD RLPEII+ M++ DLLLITADHGNDPTYAGTDHTREYIPLLA

50 Sbjct: 301 FGHRDPHGYRDCLEHFDNRLPEIIANMKEDDLLLITADHGNDPTYAGTDHTREYIPLLA 360

Query: 361 YSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDL 402
YS SFTGNGLIP GHFADISATVA+NFGVDTAMIGESFL L

55 Sbjct: 361 YSVSFTGNGLIPQGHFADISATVAENFGVDTAMIGESFLSHL 402

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2021

A DNA sequence (GBSx2132) was identified in *S.galactiae* <SEQ ID 6253> which encodes the amino acid sequence <SEQ ID 6254>. Analysis of this protein sequence reveals the following:

55 Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -12.05 Transmembrane 9 - 25 (4 - 35)

60 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2276-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6255> which encodes the amino acid sequence <SEQ ID 6256>. Analysis of this protein sequence reveals the following:

5 Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.57 Transmembrane 41 - 57 (38 - 60)
 ----- Final Results -----
 10 bacterial membrane --- Certainty=0.3230(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9143> which encodes the amino acid sequence
 15 <SEQ ID 9144>. Analysis of this protein sequence reveals the following:

 Possible cleavage site: 49
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -5.57 Transmembrane 13 - 29 (10 - 32)
 20 ----- Final Results -----
 bacterial membrane --- Certainty= 0.323(Affirmative) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 276/544 (50%), Positives = 368/544 (66%), Gaps = 5/544 (0%)

Query: 5 FKKKVVKVCLVIFGIVLVSLLSLGFFYFSKGQVLSRFVAARSRTSGQAFDNIKEYMVWSD 64
 F K +K +I L L G FY+SK ++ ++ ARS SG F+NIK Y+VW D
 30 Sbjct: 33 FHHKKLKQITIIAATSLFLFLIGGAFYYSKNHCINAYLKARSAQSGPFVFNKAYLVWDD 92
 Query: 65 TGESITNDEANYANFEPLSKSEARKLGQEIKEGNKNSMYLKRVGSRGLGIFPDYRIANKP 124
 T E ITNDEA Y F S+ E R+ Q++K +++ ++ +K VG R IFPDYRIA KP
 35 Sbjct: 93 TNEQITNDEAMYTKFRFYSQKELRQKQDLKAASQDSAVQKSVGRRFWIFPDYRIAIPK 152
 Query: 125 MSLTLKTNVPKLDVLLNQKKVATSNSDHFSVTIVERLPRTHYTASLEGTSDGKEIKLKKDY 184
 M LT+KTNVP+ DVLLNQKKVA S+S+ FSV ++RLP YTAS+ G +G+ IK+ K Y
 40 Sbjct: 153 MDLTIKTNVPQADVLLNQKKVAVSDSEQFSVKLDRLPTAETASIRGKHNGRNIKVNKSY 212
 Query: 185 DGKNQITIDLSVAFKSFTVTSNLMGDNLYFGDNRIAKLKDGSHSVNYPVTDGSKAYIKKV 244
 DG N +DLSV+F++F VTSN G+LYF DN I LKDG VE+YPVT+ ++AY+K
 45 Sbjct: 213 DGDNPVLDLSVSFRFTLVTSNAKQGDLYFDDNHIGTLKDGQLQVEDYPVTENAQAYMKTT 272
 Query: 245 FNDGEITSHKQKLISIADNQITIKLDVDGLLNEKEAGQKLITAFNQILLYVSTGQDPQTLG 304
 F.DGE+ S K L + + T+++ V LL E +AG+ L++AF+QL+ Y+STGQD L
 50 Sbjct: 273 FPDGELRSQKYALADVEEGATLEILVTDLLEEDKAGELLVSAFDQLMHYLSTGQDSSNLR 332
 Query: 305 TVFEKGAENDFYKGLKESIKAKFVTDNRKASHFTIPNIVLNKMTQVGKESYQVNFAADYD 364
 +VFE G+ N FY+GLKESIKAKF TD RKAS IP+I+L MTQVGK +Y ++F A Y+
 55 Sbjct: 333 SVFEAGSSNAFYRGLKESIKAKFQTDTRKASRLNIPSILLTMTQVGKTTYVLDFTATYE 392
 Query: 365 FNYDKSTDPDKKTYGHHIQLNLTGNFIMKSGNSYLLISNDGKKDITVAKETNKVKADPVSI 424
 F YDKSTDP++ T GHI Q+LTG +KK G YLIS G K+ITV KE N++KA S+
 60 Sbjct: 393 FLYDKSTDPQHTSGHINQDLTGKVTVKVKGQHYLLISQSGSKNITVVKEDNQLKAP--SV 450
 Query: 425 FPNLVGSGWKGEVEDGTVTMTFDKDGKVTQK-KVYKDSKSKESNHSKAVTKLEDKGNGLY 483
 FPE+++G+W G+ ++ M+ DG +T K + K ++SKE+ +AK++K+EDKGNNG Y
 65 Sbjct: 451 FPESILGTWTGQANGLSIHMSLASDGTITTKVEDQKGNRSKET-RTAKISKVEDKGNNGFY 509
 Query: 484 LYQYESGTDTTTFV-TGGIGGLKVKYAYGIKIEGNKIIPVIWQTSSDGEFDYHKPLLSKP 542
 LY + G+D + V GG+GG VKYAYG KI G PV+WQ + EFDY KPL
 70 Sbjct: 510 LYTPDPGSDISALVPEGGLGGANVKYAYGFKISGKTASPVVWQAALTHEFDYTKPLSGVT 569

-2277-

Query: 543 LTKQ 546
 L KQ
 Sbjct: 570 LQKQ 573

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9065> which encodes amino acid sequence <SEQ ID 9066>. An alignment of the GAS and GBS sequences follows:

Score = 47.3 bits (110), Expect = 4e-07
 Identities = 65/303 (21%), Positives = 119/303 (38%), Gaps = 18/303 (5%)

Query: 153 FYILGIGTSISIVVALTRFVKEISLNFKEIKKLANKMGIEVLSENENYSQII---EFDDI 209
 +YIL + T I+ +V + +S F +KKL KM + +QI EF D+
 Sbjct: 37 YYILSV-TIIACIVGGIVNLFLLSSVFTSLKKLKQKMKDISQRCFDTKAQICSPQEFKDL 95

Query: 210 LRTLHIKGDNLKSLIEREILEKQDLSFQIAALSHDIKTPXXXXXXXXXXXXXXXXXXXXQE 269
 + L+S + +++ + IA LSHDIKTP +
 Sbjct: 96 ETAFNQMSSELESTFKSLNESEREKTMIAQLSHDIKTPITSIQSTVEGILDGIISEEV 155

Query: 270 GYIVSMNNSISVFEFGYFNSLISYTRML-----SEDRSVKLILVEELLSELHFEVDDL 321
 Y + N+IS N L+ + +E + I +++LL ++ E +
 Sbjct: 156 NYYL--NTISRQTNRLNHLVEELSFITLETMSDTAEPHKEETIYLDKLLIDILSEFQLV 212

Query: 322 LNINNIEFSICNRLIITSFYGDEENLIRALSNNLLVNNAIRFMPVLDKKIEVILSESGEQIH 381
 N + I ++ + L R L NL+ NA ++ + + + I
 Sbjct: 213 FEKENRQVMIDVAPDVSKLSSQYDKLSRILLNLISNAXKYS DP-GSPLTIKAYSNRQDIV 271

Query: 382 FEIWNNGERFSDTLKKGDKLFYTEDYSRGNK--HYGIGLAFVKGVAIKHGGNLQLNPA 439
 +I + G D L Y + SR K +G+GL + +A + G++ + +
 Sbjct: 272 IDIIDQGYGIKDEDLASIFNRLYRVSSRNMTGGHGLGLYIARQLAHQLNGDILVESQY 331

Query: 440 RGG 442
 + G
 Sbjct: 332 QKG 334

A related sequence was also identified in GAS <SEQ ID 9135> which encodes the amino acid sequence <SEQ ID 9136>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -3.56 Transmembrane 145 - 161 (145 - 164)

----- Final Results -----

bacterial membrane --- Certainty=0.2423(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

SEQ ID 6254 (GBS280) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 8; MW 63.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 7; MW 88.7kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2022

A DNA sequence (GBSx2133) was identified in *S.agalactiae* <SEQ ID 6257> which encodes the amino acid sequence <SEQ ID 6258>. This protein is predicted to be ribosomal large subunit pseudouridine synthase D (rluC). Analysis of this protein sequence reveals the following:

Possible site: 22

-2278-

>>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -4.62 Transmembrane 2 - 18 (1 - 19)

----- Final Results -----

5 bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB12749 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 97/251 (38%), Positives = 147/251 (57%), Gaps = 15/251 (5%)

Query: 86 KHVLIINNEFINWQTVVQENDTITLIFDDEDYPTKKIPLGRAELIDCLYEDEHLIIVNKPE 145
 + + +N+E + +V++ D + + + + + G +D L+ED H++I+NKP

15 Sbjct: 43 QQIKVNHESVLNNMIVKKGDRVFIDLQSEASSVIPEYGE---LDILFEDNHMLIINKPA 99

Query: 146 GMKTHGNQPNIEIALLNHVSAY----SGQTCYV--VHRLDMETSGAVLFAKNPFILPLINQ 199
 G+ TH N+ + L ++ AY +G+TC V VHRLD +TSGA++FAK+ +++Q

20 Sbjct: 100 GIATHPNEDGQTGTLANLIAYHYQINGETCKVRHVHRLDQDTSGAIVFAKHRLAHAILDQ 159

Query: 200 RLERKEIWREYWALVEGKFSPKHQVLRDKIGRNR-HDRRKRIIDSKNGQHAMIIDVL-- 256
 +LE+K + R Y A+ EGK K + IGR+R H R+R+ S GQ A+T V+

25 Sbjct: 160 QLEKKTLLKRTYTAIAEGKLRTKKGTINPPIGRDRSHPTRRRV--SPGGQTAVTHFKVMAS 217

Query: 257 KYIQNSSLIKCRLETGRTHQIRVHLSHHGHPLIGDPLYNPSSN-NERLMLHAHRLTSLHP 315
 + SL++ LETGRTHQIRVHL+ GHPL GD LY S R LHA+++ HP

30 Sbjct: 218 NAKERLSLVELELETGRTHQIRVHLASLGHPLTGDSLYGGGSKLLNRQALHANKVQAVHP 277

Query: 316 LTCETISVEAP 326

30 +T E I EAP

Sbjct: 278 ITDELIVAEAP 288

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6259> which encodes the amino acid sequence <SEQ ID 6260>. Analysis of this protein sequence reveals the following:

35 Possible site: 38
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4198(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/278 (61%), Positives = 212/278 (75%), Gaps = 2/278 (0%)

45 Query: 63 TVKELLEDYFLIPRKIRHFLRVKKHVLIINNEFINWQTVVQENDTITLIFDDEDYPTKKIP 122
 TVK LLE+ LIPRKIRHFLR KKHVLIN +NWQ+ V+ D + L FD EDYP K I
 Sbjct: 2 TVKALLEEQLLIPRKIRHFLRTKKHVLIINGHSVNWQSCVKYGDQVKLFFDHEDYPEKIIIV 61

50 Query: 123 LGRAELIDCLYEDEHLIIVNKPEGMKTHGNQPNIEIALLNHVSAYSGQTCYVVHRLDMETS 182
 +G+AE + CLYEDEH+IIVNKPEGMKTHGN P E+ALLNHVSAY+GQTCYVVHRLD ETS
 Sbjct: 62 MGQAEKVTCLYEDEHIIIVNKPEGMKTHGNDPTLALLNHVSAYTGQTCYVVHRLDKETS 121

55 Query: 183 GAVLFAKNPFILPLINQRLERKEIWREYWALVEGKFSPKHQVLRDKIGRNRHRRRKRIID 242
 GA+LFAK PFILP++N+ LE+++I REY ALV G IGR+RHDRRK++D
 Sbjct: 122 GAILFAKTPFILPILNRLLLEKRDHREYALVHGSLDSPRVTYHHPIGRHRHRRRKRVVD 181

60 Query: 243 SKNGQHAMIIDVLK-YIQNSSLIKCRLETGRTHQIRVHLSHHGHPLIGDPLY-NPSSNN 300
 NG+ A+T + ++K + + +SL+ C+L+TGRTHQIRVHL+H GH L GDPLY N +
 Sbjct: 182 PINGKKAITEVTLVKNFHKRTASLLTCQLQTGRTHQIRVHLAHQGHVLFDPYLSNGKKDC 241

Query: 301 ERLMLHAHRLTSLHPLTCTETISVEAPSSTFEKILNNYK 338

RLMLHA++L L HPLT E I V+A S+TF+ +LN K

Sbjct: 242 ARLMLHAYQLRLKHPLTQEDICVQAKSATFDAVLNAQK 279

-2279-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2023

- 5 A DNA sequence (GBSx2134) was identified in *S.agalactiae* <SEQ ID 6261> which encodes the amino acid sequence <SEQ ID 6262>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.02 Transmembrane 98 - 114 (93 - 119)

----- Final Results -----

bacterial membrane --- Certainty=0.4609(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF04735 GB:AF101780 penicillin-binding protein 2a

[Streptococcus pneumoniae]

Identities = 424/773 (54%), Positives = 555/773 (70%), Gaps = 47/773 (6%)

Query: 2 KLFDFIDLFRVDEDNDEMTRKNEQETREETSNLGEEVYDIDDITRPSKSQYQGRIRHQ 61

KLF+KF+ LF+ +ETS L+ + I R S+S

Sbjct: 5 KLFKFLSLFK-----KETSELEDSD----STILRRSRS----- 34

Query: 62 KENAKSRPEWLQKVDRLPSKPNPIRRFWRRYRIGKLLFIALMAFILIFGSYLFYLSKTA 121

DR + PIR+FWRRY + K++ I ++ L+ G YLF ++K+

Sbjct: 35 -----DRKKLAQVGPIRKFWRRYHLTKIILILGLSAGLLVGIYLFYFAVAKST 80

Query: 122 TVSDLQSALKTTTTIYDKNKEYAGKLSGQKGTVELNAISDHLKNAVIATEDRTFYENNG 181

V+DLQ+ALKT T I+D+ ++ AG LSGQKGTVEL IS +L+NAVIATEDR+FY+N+G

Sbjct: 81 NVNDLQNALKTRTLIFDREEKEAGALSGQKGTVELTDISKNLQNAVIATEDRSFYKNDG 140

Query: 182 VNFKRFFLAVATLGKFGGSGTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSKAEI 241

+N+ RFFLA+ T G+ GGGSTITQQLAKNAYLSQDQ++RKA+EFFLALEL+KKYSK +I

Sbjct: 141 INYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVKAKEFFLALELSKKYSKEQI 200

Query: 242 LTMYLNNSYFNGVWGVEDASRKYFGTSAANLTVDEAATLAGMLKGPEVYNPYYSVENAT 301

LTMYLNN+YFNGVWGVEDAS+KYFG SA+ +++D+AATLAGMLKGPE+YNP SVE++T

Sbjct: 201 LTMYLNNAYFNGVWGVEDASKKYFGVSASEVSLDQAATLAGMLKGPELYNPLNSVEDST 260

Query: 302 NRRDITVLAAMVDAGKLTKSQAKEAASIGMKNRLADTYAGKINDYRPSYFDVAVNEAIDT 361

NRRDITVL MV AG + K+Q EAA + M ++L D Y GKI+DYRPSYFDVAVNEA+

Sbjct: 261 NRRDITVLQNMVAAGYIDKNQETEAEEVDMTSQLHDKYEGKISDYRPSYFDVAVNEAVSK 320

Query: 362 YGISEKDIVNNGYKIY TALDQNYQSGMQKTFDDTSLFPVSDYDQSAQGASVALDPKTGG 421

Y ++E++IVNNGY+IYT LDQNYQ+ MQ +++TSLFP ++ DG AQ SVAL+PKTGG

Sbjct: 321 YNLTEEEIVNNGYRIY TELDQNYQANMQIVYENTSLFPRAE-DGTFAQSGSVALEPKTGG 379

Query: 422 VRGLVGRVQSTKDAQFRSFNYATQSKRSPASTIKPLVVYSPAIASGWSIDKELPNKVQDF 481

VRG+VG+V FR+FNATQSKRSP STIKPLVVY+PA+ +GW+++K+L N +

Sbjct: 380 VRGVGVQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPAVEAGWALNKQLDNHTMQY 439

Query: 482 HGYKPSNYGGIET-ESIPMYQALANSYNIPAVYTLDKLGINKAFTYGRKFGFLNMSSANKE 540

YK NY GI+T +PMYQ+LA S N+PAV T++ LG++KAF G KFGFLNM ++

Sbjct: 440 DSYKVDNYAGIKTSREVPMYQSLAESLNLPAVATVNDLGVDKAFEAGEKFGFLNMEKVDRV 499

Query: 541 LGVALGGSVTINPLEMAQAYSTFANDGIMHRAHLITRIETANGKLVKQFTDKPKRVISRS 600

LGVALG V TNPL+MAQAY+ FAN+G+M AH I+RIE A+G+++ + KRVI +S

Sbjct: 500 LGVALGSGVETNPLQMAQAYAAAFANEGLMPEAHFISRIENASGQVIASHKNSQKRVIDKS 559

Query: 601 VASKMTSMMLGTFSNGTAINANVGYTMMAGKTGTETDFNPNLSGDQWVVGYPDPVVISQ 660

VA KMTSMMLGT+NGT I+++ Y MAGKTGTTE FNP + DQWV+GYTPDVVIS

-2280-

Sbjct: 560 VADKMTSMMLGTFTNGTGISSSPADYVMAGKTGTTEAVFNPEYTSQWVIGYTPDVVISH 619

Query: 661 WVGFKNTDKHHYLTDSAGTASNIFSTQASYILPYTKGSSFTHIENAYFQNGIGSVYNAQ 720
W+GF TD++HYL S++ A+++F A+ ILPYT GS+FT +ENAY QNGI +

5 Sbjct: 620 WLGFPTTDENHYLAGSTSNGAHVFRNIANTILPYTPGSTFT-VENAYKQNGIAPANTKR 678

Query: 721 DASNTTNQESRSIINDLKDSASKAAQDISRAVEDSNFQEKVKDAWNSLKDYFR 773
N ++ ++D++ A + SRA+ D+ +EK + W+S+ + FR

10 Sbjct: 679 QVQTNDNSQTDDNLSDIRGRAQSLVDEASRAISDAKIKEKAQTIWDSIVNLFR 731

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6263> which encodes the amino acid sequence <SEQ ID 6264>. Analysis of this protein sequence reveals the following:

Possible site: 52

15 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -7.96 Transmembrane 104 - 120 (99 - 124)

----- Final Results -----

20 bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:AAF04735 GB:AF101780 penicillin-binding protein 2a [Streptococcus pneumoniae]
Identities = 414/730 (56%), Positives = 539/730 (73%), Gaps = 17/730 (2%)

30 Query: 50 TKNSEQDPATALQRSRAYEGSPKSRPAWLQKLEAVLPSPQRPPIRRFWRRYHIGKLLMILI 109
T E +T L+RSR+ +KL V PIR+FWRRYH+ K+++IL
Sbjct: 18 TSELESDSTILRRSRSDR-----KKLAQV-----GPIRKFWRRYHILTKIILILG 62

Query: 110 GTLVLLLGSYLFYLSKTAKVSDLDQALKATTVIYDHKGEYAGSLSGQKGSYVELNAISDD 169
+ LL+G YLF ++K+ V+DLQ+ALK T+I+D + + AG+LSGQKG+YVEL IS +
Sbjct: 63 LSAGLLVGIYLFVAVAKSTNVNDLQNALKTRTLIFDREEKEAGALSQKGTVELTDISK 122

35 Query: 170 LENA VIATEDRTFYNSGINKRFLAVVTAGRFGGGGSTITQQLAKNAYLSQDQTIKRKA 229
L+NAVIATEDR+FY N GIN RF LA+VTAGR GGGSTITQQLAKNAYLSQDQT++RKA
Sbjct: 123 LQNAVIATEDRSFYKNDGINYGRFFLAIVTAGRSGGGSTITQQLAKNAYLSQDQTVRKA 182

40 Query: 230 REFFLALELTKKYSKKDILTYMLNNSYFGNGVWGVEDASQKYFGTTAANLTLDEAATLAG 289
+EFFLALEL+KKYSK+ ILTYMLNN+YFGNGVWGVEDAS+KYFG +A+ ++LD+AATLAG
Sbjct: 183 KEFFLALELSKKYSKEQILTMYLNAYFGNGVWGVEDASKKYFGVSASEVSLDQAATLAG 242

45 Query: 290 MLKGPEIYNPYHSLKNATHRRDITVLGAMVDAKKITQTKAQARAVGLKNRLADTYVGKTD 349
MLKGPE+YNP +S++++T+RRDITVL MV A I + + +A V + ++L D Y GK
Sbjct: 243 MLKGPELYNPLNSVEDSTNRRDITVLQNMVAAGYIDKNQETEAEEVDMTSQLHDKYEGKIS 302

50 Query: 350 DYKYPYSYFDAVISEAIATYGLSEKDIVNNGYKVYTELDQNYQTGMQTTFNDELFPVSAY 409
DY+YPSYFDAV++EA++ Y L+E++IVNNGY++YTELDQNYQ MQ + N LFP A
Sbjct: 303 DYRYPYSYFDAVNEAVSKYNLTETEEIVNNGYRIYTELDQNYQANMQIVYENTSLFP-RAE 361

Query: 410 DGSSAQAAASVALDPKTGGVRLIGRVNSENPTFRSFNYATQAKRSPASTIKPLVVYAPA 469
DG+ AQ+ SVAL+PKTGGVVG++G+V ++ FR+FNYATQ+KRSP STIKPLVVY PA
Sbjct: 362 DGTFAQSGSVALEPKTGGVVGQVADNDKTGFRNFNYATQSKRSPGSTIKPLVVYTPA 421

55 Query: 470 VASGWSIEKELPNTVQDFDGYQPHNY-GNYESEDVPMYQALANSYNIPAVSTLNDIGIDK 528
V +GW++ K+L N +D Y+ NY G S +VPMYQ+LA S N+PAV+T+ND+G+DK
Sbjct: 422 VEAGWALNKQLDNHTMQYDSYKVDNYAGIKTSREVPMYQSLAESLNLPAVATVNDLGVDK 481

60 Query: 529 AFTYKGTFLGLDMSSAKKELGVALGGSVTTNPLEMAQAYAAAFANNGVIHPAHLINRIENAR 588
AF G+ FGL+M + LGVALG V TNPL+MAQAYAAAFAN G++ AH I+RIENA
Sbjct: 482 AFEAGEKFGFLNMEKVDRVLGVALGSGVETNPLQMAQAYAAAFANGLMPEAHFISRIENAS 541

65 Query: 589 GEVLKTFITDKAKRVVSQSADKMTAMMLGTFSNGTAVNANVYGYTLAGKTGTTETNFNPD 648
G+V+ + + KRV+ +SVADKMT+MMLGT+NGT +++++ Y +AGKTGTTE FNP+
Sbjct: 542 GQVIASHKNSQKRVIDKSVADKMTSMMLGTFTNGTGISSSPADYVMAGKTGTTEAVFNPE 601

-2281-

Query: 649 LAGDQWVIGYTPDVVISQWVGFNQTDENHYLTDSSAGTASAIIFSTQASYILPYTKGSQFH 708
 DQWVIGYTPDVVIS W+GF TDENHYL S++ A+ +F A+ ILPYT GS F
 Sbjct: 602 YTSDQWVIGYTPDVVISHWLGFPTDENHYLAGSTSNAAHVFRNIANTILPYTPGSTFT 661

Query: 709 VDNAYAQNGISAVYGVNETGNQSGVDTQSIIDGLRKSAQEASQSLSKAVDQSGLRDKAQS 768
 V+NAY QNGI+ + T + +R AQ S+A+ + +++KAQ+
 Sbjct: 662 VENAYKQNGIAPANTKRQVQTNDNSQTDNLSDIRGRAQSLVDEASRAISDAKIKEKAQT 721

Query: 769 IWKEIVDYFR 778
 IW IV+ FR
 Sbjct: 722 IWDSIVNLF 731

An alignment of the GAS and GBS proteins is shown below.

Identities = 530/715 (74%), Positives = 623/715 (87%), Gaps = 1/715 (0%)

Query: 59 RHQKENAKSRPEWLQKVDRLPSPKNPIRRFWRRYRIGKLLFIALMAFILIFGSYLFYLS 118
 R + + KSRP WLQK++ LPSP+ PIRRFWRRY IGKLL I + +L+ GSYLFYLS
 Sbjct: 65 RAYEGSPKSRPAWLQKLEAVLPSPQRPPIRRFWRRYHIGKLLMILIGTLVLLLGSYLFYLS 124

Query: 119 KTATVSDLQSAKLTITTIYDKNKEYAGKLSGQKGYVELNAISDHLKNAVIATEDRTFY 178
 KTA VSDLQ ALK TT IYD EYAG LSGQKG+YVELNAISD L+NAVIATEDRTFY
 Sbjct: 125 KTAKVSDLQDALKATTVIYDHKGEYAGSLSGQKGSYVELNAISDDLNAVIATEDRTFY 184

Query: 179 NNGVNFKRFFLAVATLGKFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK 238
 N+G+N KRF LAV T G+FGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK
 Sbjct: 185 NSGINLKRFLAVVTAGRFGGGSTITQQLAKNAYLSQDQTIKRKAREFFLALELTKKYSK 244

Query: 239 AEILTMYLNNISYFGNGVWGVEDASRKYFGTSAANLTVDEAATLAGMLKGPEVYNPYYSVE 298
 +ILTMYLNNISYFGNGVWGVEDAS+KYFGT+AAANLT+DEAATLAGMLKGPE+YNPY+S++
 Sbjct: 245 KDILTMYLNNISYFGNGVWGVEDASQKYFGTTAANLTLDEAATLAGMLKGPEIYNPYHSLK 304

Query: 299 NATNRRDVTLAAMVDAGKLTQSQAKEAASIGMKNRLADTYAGKINDYRPSYFDAVVNEA 358
 NAT+RRDVTL AMVDA K+T+++A+++A ++G+KNRLADTY GK +DY+YPSYFDAV++EA
 Sbjct: 305 NATHRRDVTLGAMVDAAKKTQTQKQARAVGLKNRLADTYVGKTDYKPSYFDAVISEA 364

Query: 359 IDTYGISEKDIVNNGYKIYALDQNYQSGMQKTFDDTSLFPVSDYDQSAQAGSVALDPK 418
 I TYG+SEKDIVNNGYK+YT LDQNYQ+GMQ TF++ LFPVS YDG SAQ ASVALDPK
 Sbjct: 365 IATYGLSEKDIVNNGYKVYTELDQNYQTGMQTTFNDELFPVSAYDGSSAQAASVALDPK 424

Query: 419 TGGVRLVGRVQSTKDAQFRSFNYATQSKRSPASTIKPLVVYSPAISGWSIDKELPNKV 478
 TGGVRLG+GRV S+++ FRSFNYATQ+KRSPASTIKPLVVY+PA+ASGWSI+KELPN V
 Sbjct: 425 TGGVRLIGRVNSENPTFRSFNYATQAKRSPASTIKPLVVYAPAVASGWSIEKELPNTV 484

Query: 479 QDFHGYKPSNYGGIETESIPMYQALANSYNIPAVYTLDKLGINKAFTYGRKFGLNMSSAN 538
 QDF GY+P NYG E+E +PMYQALANSYNIPAV TL+ +GI+KAFTY+ FGL+MSSA
 Sbjct: 485 QDFDGYQPHNYGNYESEDVPMYQALANSYNIPAVSTLNDIGIDKAFTYKTFGLDMSSAK 544

Query: 539 KELGVALGGSVTTNPLEMAQAYSTFANDGIMHRAHLITRIETANGKLVKQFTDKPKRVIS 598
 KELGVALGGSVTTNPLEMAQAY+ FAN+G+++H AHLI RIE A G+++K FTDK KRV+S
 Sbjct: 545 KELGVALGGSVTTNPLEMAQAYAAAFANNGVIHPAHLINRIENARGEVLKFTFDKAKRVVS 604

Query: 599 RSVASKMTSMMLGTFSNGTAINANVYGYTMAGKTGTTETDFNPNLSGDQWVVGYPDVVI 658
 +SVA KMT+MMLGTFSNGTA+NANVYGYT+AGKTGTTET+FNPN+L+GDQWV+GYTPDVVI
 Sbjct: 605 QSVADKMTAMMLGTFSNGTAVNANVYGYTLAGKTGTTETNFPNPDLAGDQWVIGYTPDVVI 664

Query: 659 SQWVGFKNTDKHHYLTDSAGTASNIIFSTQASYILPYTKGSSFTHIENAYFQNGIGSVYN 718
 SQWVG F TD++HYLTDSAGTAS IFSTQASYILPYTKGS F H++NAY QNGI +VY
 Sbjct: 665 SQWVGFNQTDENHYLTDSSAGTASAIIFSTQASYILPYTKGSQF-HVDNAYAQNGISAVYG 723

Query: 719 AQDASNTTNQESRSIINDLKDSASKAAQDISRAVEDSNFQEKVKDAWNLSLKDYFR 773
 + N + +++SII+ L+ SA +A+Q +S+AV+ S ++K + W + DYFR
 Sbjct: 724 VNETGNQSGVDTQSIIDGLRKSAQEASQSLSKAVDQSGLRDKAQSIIWKEIVDYFR 778

SEQ ID 6262 (GBS397d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 153 (lane 13; MW 76kDa) and in Figure 184 (lane 9; MW 76kDa).

A DNA sequence (GBSx2135) was identified in *S.agalactiae* <SEQ ID 6265> which encodes the amino acid sequence <SEQ ID 6266>. This protein is predicted to be M-like protein. Analysis of this protein sequence reveals the following:

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS gene <SEQ ID 8957> and protein <SEQ ID 8958> were also identified. Analysis of this protein sequence reveals the following:

LPXTG motif: 596-600

```

      282          312          342          372          402          432          462          666
LS**IRIFN*LYKGANMNNNEKKVKYFLRKTAYGLASMSAAFVCSGIVHADTSSGISASIPHKQVNLGAVT~~~THA
      :|::||:||:||||:||||:| |: | ||:
MAKKEMKFYLRLKSAFGLASVSAAALLVGAARVSADS-----~~~~~
              10              20              30

```

921 951 981 1011 1041 1071 1101 1131
PEV---LEEYKEKIQRASTKQVDFVAEAKKVVNSNKETLVNQANGKKQEI~~AKLENLSNDEMLRYNTAIDNVVKQYN~~EG
|: : | ::||| | : :| :: : :| | | : : | : |
AELEQKIPELEKKIQEAQEKLNYENRPS-~~PKRVGSDDEDDTVARKLMSEKAL~~KA~~E-----~~LQKTKEALDTAKRAYAGI
110 120 130 140 150 160 170

1161 1191 1221 1251 1281 1311 1638 1668
KLNITAAMNALNSIKQAAQVEVAQNLOQYAKKIERISSKGLALSKKAKEIYEKHKSILPTP~~~~~AKPDVKPEAKPDVK
:
| | :: : | | : | | : |
EERKQVAATKLDAANKAFAGVEEKHAQAMAFAAFAAYKGA-----

 190 200 210

1698 1740 1770 1800 1830
PKAKPDVKPEAKPDVKPD-----VKPDVKPEAKPEDKPDVKPDVKPEAKPDVKPEAKPE
||| | |||:||||| : ||: ||| ||: ||| |||
-----VKAEKKAAGASDFYTKKIDSADTVDGVKTLREMILDSIAKPEVEPEAKPEPKLEPKPEPKPEPKPEPKPE
220 230 240 250 260 270 280

1860 1890 1920 1950 1980 2010 2040 2070
AKPEAKPEAKPEAKPEAKPDVKPEAKPDVKPEAKPEAKPEAKSEAKPEAKLEAKPEAKPATKKSVENTSGNLAAKKAIENTK
||| ||| ||| ||| ||: ||: :| | |||| | |
PKPEPKPEPKPEPKPEPKPKPKQPKPAAPKPEAKKEEKKAAP-----K
300 310 320 330

2100 2130 2160 2190 2220 2250 2280 2310
KYSKKLPSTGEAASPLLAIVSLIVMLSAGLITIVLKHKKN*IYF*T*TERSILSKS*GKPHQNFAPFI*ILE*FSRYFN*
: : ||||| : :: : || ||: : : |:
QDTNKLPSSTGEATNPFFTAALAVMAGAGVAAVSTRKEN
350 360 370

SEQ ID 6266 (GBS3) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 5; MW 65kDa). The GBS3-His fusion product was purified (Figure 189, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 261), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2284-

Example 2025

A DNA sequence (GBSx2136) was identified in *S.agalactiae* <SEQ ID 6267> which encodes the amino acid sequence <SEQ ID 6268>. This protein is predicted to be transcription antitermination protein nusg (nusG). Analysis of this protein sequence reveals the following:

```

5   Possible site: 48
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3203(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAA53738 GB:X76134 nusG [Staphylococcus carnosus]
    Identities = 90/175 (51%), Positives = 118/175 (67%), Gaps = 2/175 (1%)

    Query: 7   KGWFLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGKTKEIENRFP 66
              K W+ + TYSGYENKVK+NL +R ++ NM + I RV IP +      K+GK K++ + FP
    Sbjct: 8   KRWYAVHTYSGYENKVKKNLEKRVESMNMTEQIFRVVPIPEEETQVKDGAKKLTKKTFP 67

20   Query: 67  GYVLVEMVMTDEAWFVVRNTPNVTGFVSGHGNRSKPTPLLEEIRSIILISMGTVDVFD 126
              GYVLVE+VMTDE+W+VVRNTP VTGFVGS G SKP PLL +E+R IL MG      D
    Sbjct: 68  GYVLVELVMTDESWEVVRNTPGVTFVGSAGAGSKPNPLLPDEVRFILKQMGMEKKTIDV 127

25   Query: 127 NIKEGDVVQIIDGAFIQEGRVVEIENNKVKL--MINMFGSETQAELELYQVAEL 179
              ++ G+ V+I  G F  Q G V EIE +K KL  +++MFG ET  E+E  Q+ +L
    Sbjct: 128 EVEVGQVRIKSGPFANQVGEVQEIEADKFKLTLVDMFGRETPEVEFEQIEKL 182

```

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6269> which encodes the amino acid sequence <SEQ ID 6270>. Analysis of this protein sequence reveals the following:

```

    Possible site: 53
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
35      bacterial cytoplasm --- Certainty=0.3874(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40   Identities = 170/179 (94%), Positives = 178/179 (98%)

    Query: 1   MLDSFDKGFVFLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGKTKEI 60
              MLDSFDKGFVFLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNG+TKEI
    Sbjct: 6   MLDSFDKGFVFLQTYSGYENKVKENLLQRAQTYNMLDNILRVEIPTQTVNVEKNGQTKEI 65

45   Query: 61  EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFVSGHGNRSKPTPLLEEIRSIILISMGT 120
              EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFVSGHGNRSKPTPLLEEIR+IL+SMGT
    Sbjct: 66  EENRFPGYVLVEMVMTDEAWFVVRNTPNVTGFVSGHGNRSKPTPLLEEIRAILLSMGT 125

50   Query: 121 VDVFDNIKEGDVVQIIDGAFIQEGRVVEIENNKVKLMINMFGSETQAELELYQVAEL 179
              +DVFDNIKEGDVVQIIDGAF+QEGRVVEIENNKVKLM+NMFGSET AE+ELYQ+AE
    Sbjct: 126 IDVFDNIKEGDVVQIIDGAFMQEGRVVEIENNKVKLMINMFGSETVAEVELYQIAEL 184

```

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2285-

Example 2026

A DNA sequence (GBSx2137) was identified in *S.agalactiae* <SEQ ID 6271> which encodes the amino acid sequence <SEQ ID 6272>. This protein is predicted to be a glycosyl transferase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 16
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1558(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
    ducreyi]
    Identities = 98/259 (37%), Positives = 155/259 (59%), Gaps = 10/259 (3%)

    Query: 5   VALAVDSNYLDKALVTIKSICVYNRNITFYLFNQDTPVEWVRNINRKLEPLGSKLINVKI 64
      + LA + +Y + L TIKSI ++N++I FYL N+D P EW +N KL L S++I++K+
20  Sbjet: 10  IVLAANQSYSEYILTTFIKSIYLNKHIRFYLLNRDYPTEWFDILNNKLRLKLNSEIIDIKV 69

    Query: 65  YNYDIAHLTTFLTVS---TWFRFLADYIPSSRVLYLDSDIIVNTNLDYLFELDFKGYL 121
      N I + T+ +S T+FR F++D+I +V+YLD+DI+VN +L L++ D Y+L
25  Sbjet: 70  TNDTIKNFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISNYFL 129

    Query: 122 AAVKDPHKNE----EGGFNAGMLLANLELWREDGLTKTLLKTAEELHRVVKTGDQSILNI 177
      AAVKD + FNAGMLL N + WRE +T+ L +E+ + DQSILN+
30  Sbjet: 130 AAVKDIISEKIYVNNHIFNAGMLLINNKKWREHNITQFCLSLSEKYINSLPDADQSILNL 189

    Query: 178 VCHNRWLSLNKTWNF--QTYDVVSRYNHRSYLYLNENRTPNIIHFLTSDKPWNENSVAR 235
      + ++WL LN+ +N+ T + +Y YL ++ P IIH+ T KPW R
35  Sbjet: 190 IFKDKWLKLNRGYNYLIGTDYLFYFKYKTRYLE-DLGETIPLIIHYNTEAKPWLNIFNTR 248

    Query: 236 FRELWWYYFQLDFCQLTGK 254
      FR ++W+Y++L++ + K
40  Sbjet: 249 FRNIYWFFYELNWQDIYAK 267

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2027

A DNA sequence (GBSx2138) was identified in *S.agalactiae* <SEQ ID 6273> which encodes the amino acid sequence <SEQ ID 6274>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 20
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
50      bacterial cytoplasm --- Certainty=0.0417(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2286-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2028

A DNA sequence (GBSx2139) was identified in *S.agalactiae* <SEQ ID 6275> which encodes the amino acid sequence <SEQ ID 6276>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -2.60    Transmembrane    306 - 322 ( 306 - 322)

----- Final Results -----
          bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF28363 GB:AF224467 putative glycosyl transferase [Haemophilus
    ducreyi]
    Identities = 88/259 (33%), Positives = 156/259 (59%), Gaps = 11/259 (4%)

Query: 7   VVLADYSYIRQIETTTLKSLCVYHENLSIFIFNQDIPQEWFLAMKDRVGTGNQIQDVKL 66
          +VLA + SY   I TT+KS+ ++++++ ++ N+D P EWF + +++ + ++I D+K+
Sbjct: 10  IVLAANQSYSEYIITTIKSIYLNKHIRFYLLNRDYPTEWFDILNNKLRKLNSEIIDIKV 69

Query: 67  FHDHLSPKWENKKLNHNINY-MTYARYFIPQYISADTVLYLSDLVVTNTLNDNLFQISLDN 125
          +D +      K +HI+  T+ RYFI +I D V+YLD+D+VV +L L+Q + N
Sbjct: 70  TNDTIK---NFKTYSHISSDTTFFRYFISDFIEQDKVIYLDADIVVNGSLTELYQTDISN 126

Query: 126 AYLAAVP-----ALFGLGYGFNAGVMVINNRWRQENMTIKLIEKNQKEIENANEGDQTI 180
          +LAAV      ++   + FNAG+++INN++WR+ N+T   + ++K I + + DQ+I
Sbjct: 127 YFLAAVKDIISEKIYVNNHIFNAGMLLINNKWREHNITQFCLSLSEKYINSLPDADQSI 186

Query: 181 LNRMFENQVIYLDPTYNFQIGFD-MGAAIDGHKFIFDIPITPLPKIIHYISGIKPWQTLS 239
          LN +F+++ + L+  YN+ IG D +      +++ D+  T +P I IHY + KPW +
Sbjct: 187 LNLIFKDKWLKLNRGYNYLIGTDYLFYKYGKTRYLEDLGET-IPLIIHYNTEAKPWLNIF 245

Query: 240 NMRLREVWWHYNNLEWSSI 258
          N R R ++W Y L W I
Sbjct: 246 NTRFRNIYWFYELNWQDI 264
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6276 (GBS395) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 5; MW 47.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 8; MW 72kDa) and in Figure 177 (lane 5; MW 72kDa).

GBS395-GST was purified as shown in Figure 217, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2029

A DNA sequence (GBSx2140) was identified in *S.agalactiae* <SEQ ID 6277> which encodes the amino acid sequence <SEQ ID 6278>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
```


-2287-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1633(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2030

A DNA sequence (GBSx2141) was identified in *S.agalactiae* <SEQ ID 6279> which encodes the amino acid sequence <SEQ ID 6280>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 36 - 52 (36 - 52)

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10243> which encodes amino acid sequence <SEQ ID 10244> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77330 GB:AE000508 orf, hypothetical protein [Escherichia coli K12]
 Identities = 75/260 (28%), Positives = 123/260 (46%), Gaps = 22/260 (8%)

Query: 6 VGLVLEGGGMRLGLYTAGVLDALFDAGIK-IDGIVSVSAGALFGVNFVSRQRERLRYNKK 64
 + LV EGGG RG++TAGVLD F+ A D + SAGA F+ Q A + +
 Sbjct: 25 IALVCEGGGQRGIFTAGVLDEFMRAQFNPFDLVLTGTSAGAQNLSAFICNPGYARKVIMR 84

Query: 65 YLSHPKYMRLRSWFRTGNFVNKDF----TYEVPKLD----VFDDEAFKKSSIDFYVVA 116
 Y + ++ + R GN ++ D+ T ++P+++D +FD S FY+ A
 Sbjct: 85 YTTKREFDPLRFVRGGNLIDLWLVEATASQMPLQMDTAARLFD-----SGKSFYMCA 138

Query: 117 TEMTSGKPEYFKIDSVFEQMEILRASSALPVVSKM-VDWQGGKYLDGGLSDSIPVDFARG 175
 P YF + + ++++RASSA+P + V +G YLDGG+SD+IPV A
 Sbjct: 139 CRQDDYAPNYF-LPTKQNWLDVIRASSAIPGFYRSGVSLEGINYLDGGISDAIPVKEAAR 197

Query: 176 LGFDKLIVVMTRPLNYQKKPSSGR-----LYKTLRKYPNFKVTASNRYQQYNNSLEKVM 230
 G L+V+ T P P + L + + N V+ Y+ +EK
 Sbjct: 198 QGAKTLVVIRTVPQMYTTPQWFKRMRWLGDSSLQPLVNLVQHHTSYRDIQQFIEKPP 257

Query: 231 SLEKTGDLFAIRPSKSLVIG 250
 + +++ +P S+ +G
 Sbjct: 258 GKLRIFEIYPPKPLHSIALG 277

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8959> and protein <SEQ ID 8960> were also identified. Analysis of this protein sequence reveals the following:

-2288-

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -5.16
 GvH: Signal Score (-7.5): -2.17
 Possible site: 44

>>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -0.16 threshold: 0.0
 INTEGRAL Likelihood = -0.16 Transmembrane 36 - 52 (36 - 52)
 PERIPHERAL Likelihood = 4.14 18
 modified ALOM score: 0.53

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01611(316 - 1050 of 1449)
 OMNI|NT01EC5264(37 - 289 of 369) hypothetical protein
 %Match = 9.2
 %Identity = 29.7 %Similarity = 49.8
 Matches = 74 Mismatches = 118 Conservative Sub.s = 50

273 303 333 363 393 420 450 480
 QKKQLYFAIL*SNINIRK*LPMLSVGLVLEGGMRGLYTAGVLDAFLDAGIK-IDGIVSVSAGALFGVNFVSRQRERALR
 : || |||| ||::||| ||: | | : ||| |: | | :
 VGQRIPVTLGNIAPLSLRPFQPGRIALVCEGGGQRGIFTAGVLDEFMRAQFNPFDLYLGTSAQAQNLSAFICNQPGYARK
 30 40 50 60 70 80 90

510 540 588 618 648 678 708
 YNKKYLSPKYMSLRFSWFRITGNFVNKDF----TYEVPKMLDVFDEAFKSSIDFYVVATEMTSGKPEYFKIDSVFQEM
 : | : : : : | ||::: | : | : : : : | : | | : | | | : : :
 VIMRYTTKREFFDPLRFVRGGNLDLDLWLVEATASQMLQMDT--AARLFDSGKSFYMCACRQDDYAPNYF-LPTKQNLWL
 110 120 130 140 150 160

738 765 795 825 855 885 912 930
 EILRASSALEPVVSKM-VDWQGGKYLDGGLSDSIPVDFARGLGFDKLIIVMTRPLNYQKKPSS-GRLYKTL----YRKYPN
 ::|||:| : | : | | |||:|:| || | | : : | | | : : | :
 DVIRASSAIPGFYRSGVSLIGINYLDGGISDAIPVKEAARQGAKTLVIRTVPVPSQMYTTPQWFKRMERWLGDSSLQPLVN
 180 190 200 210 220 230 240

960 990 1020 1050 1080 1110 1140 1170
 FVKTASNRYQQYNNLSLEKVMSEKTDGLFAIRPSKSLVIGRLEKNPDKLSIYQLGMKDAKSVMPELNSYLMK*RKQYFS
 : | : | : : || : : : : | | : : |
 LVQHHETSYRDIQQFIEKPPGKLRIFEIYPPKPLHSIALGSRIPALREDYKLGRLCGRYFLATVGKLLTEKAPLTRHLVP
 260 270 280 290 300 310 320

SEQ ID 8960 (GBS394) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 4; MW 34.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 7; MW 60kDa).

GBS394-GST was purified as shown in Figure 217, lane 6.

Example 2031

A DNA sequence (GBSx2142) was identified in *S.agalactiae* <SEQ ID 6281> which encodes the amino acid sequence <SEQ ID 6282>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2289-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2032

A DNA sequence (GBSx2143) was identified in *S.agalactiae* <SEQ ID 6283> which encodes the amino acid sequence <SEQ ID 6284>. This protein is predicted to be transporter protein. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have a cleavable N-term signal seq.

15	INTEGRAL	Likelihood = -6.85	Transmembrane	373 - 389 (370 - 395)
	INTEGRAL	Likelihood = -6.74	Transmembrane	168 - 184 (162 - 187)
	INTEGRAL	Likelihood = -6.32	Transmembrane	259 - 275 (257 - 280)
	INTEGRAL	Likelihood = -4.78	Transmembrane	286 - 302 (285 - 306)
	INTEGRAL	Likelihood = -3.19	Transmembrane	55 - 71 (54 - 71)
	INTEGRAL	Likelihood = -2.97	Transmembrane	84 - 100 (79 - 101)
20	INTEGRAL	Likelihood = -2.87	Transmembrane	311 - 327 (310 - 328)
	INTEGRAL	Likelihood = -1.44	Transmembrane	355 - 371 (355 - 371)
	INTEGRAL	Likelihood = -0.64	Transmembrane	108 - 124 (108 - 125)

----- Final Results -----

25	bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22759 GB:U32790 transporter protein [Haemophilus influenzae Rd]

Identities = 139/391 (35%), Positives = 221/391 (55%), Gaps = 4/391 (1%)

Query: 6 INKNNWRALIAAIVASGTTDDLNMFLAFMSSTIITDLHLSAAQAGWIGTITNLGMLVGGGL 65
+N W-ALI + V G D +++ L F +S I DL+L+ AQ G + T T +G + GG+

35 Sbjct: 5 VNSYGWKALIGSAVGYGMDGFDLLILGFMLSAISADLNLTPAQGGSILVTWTLIGAVFGGI 64

Query: 66 IFGLLADRYNKFVKVFKWTILIFSIATGLVFFTTNLSYLYIMRFIAGIGVGGEYGI AIAIM 125
+FG L+D+Y + +V WTIL+F++ TGL L I R IAGIG+GGE+GI +A+

40 Sbjct: 65 LFGALSDKYGRVRVLTWTILLFAVFTGLCAIAQGYWDLIIYRTIAGIGLGGEFGIGMALA 124

Query: 126 AGIVPTNMKGRISLNGIAGQVGSISSALLAGWLAPALGWRGLFLFGLLPITVLVLMQFA 185
A P + +S + QVG + +ALL L P +GWRG+FL G+ P + +++

Sbjct: 125 AEAAPARHRAKAASYVALGWQVCVLGAALLTPLLPHIGWRGMFLVGIFPAFVAWFLRSH 184

45 Query: 186 VDDKDILDQYNTDADDEPLDI----SIKALFDTPLVATQSLALMVMTTVQIAGYFGMMNW 241
+ + +I Q T + S + L + SL ++V+T+VQ GY+G+M W

Sbjct: 185 LHEPEIFTQKQTALSTQSSFTDKLRSEFQLLIKDKATSKISLGIVVLTSVQNFQGYGIMIW 244

Query: 242 LPTIIQTNLNVSVKNSSLWMIATILGMCLGMLVFGQLLDKFGPRLVYGCFLSSAICVYL 301
LP + L S+ S LW T+ GM G+ +FGQL D+ G + + F L + I + +

50 Sbjct: 245 LPNFLSKQLGFSLTKSLWTAVTVCGMMAGIWIIFGQLADRIGRKPSFLFLQLGAVISIVV 304

Query: 302 FQFATTMPMSMIIGGAVVGFFVNGMFAGYGAMITRLYPHIRSTANNLILNVGRAIGGFSS 361
+ T M++ GA +G FVNGM GYGA++ YP R+TA N++ N+GRA+GGF

55 Sbjct: 305 YSQLTDPDIMLLAGAFGLGMFVNGMLGGYGALMAEAYPTARATAQNVLFNIGRAVGGFGP 364

Query: 362 VIIGMILDVSNVSMVMLFLASLYIVSFLSML 392
V++G ++ + + LA +Y++ L+ +

Sbjct: 365 VVVGSVVLAYSFQTAIALLAIIYVIDMLATI 395

-2290-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2377> which encodes the amino acid sequence <SEQ ID 2378>. Analysis of this protein sequence reveals the following:

Possible site: 39

```

5  >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -8.92    Transmembrane  168 - 184 ( 162 - 188)
    INTEGRAL    Likelihood = -5.41    Transmembrane  286 - 302 ( 285 - 306)
    INTEGRAL    Likelihood = -5.15    Transmembrane  372 - 388 ( 362 - 394)
10  INTEGRAL    Likelihood = -3.45    Transmembrane  259 - 275 ( 257 - 276)
    INTEGRAL    Likelihood = -2.87    Transmembrane  311 - 327 ( 306 - 328)
    INTEGRAL    Likelihood = -2.81    Transmembrane   55 -  71 (  51 -  71)
    INTEGRAL    Likelihood = -0.48    Transmembrane  108 - 124 ( 108 - 125)
    INTEGRAL    Likelihood = -0.37    Transmembrane   84 - 100 (  84 - 100)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 306/402 (76%), Positives = 354/402 (87%)

```

Query: 1  MSPLNINKNNWRALIAAIVASGTDDL NIMFLAFSMTIITDLHLSAAQAGWIGTITNLGM 60
25  MS L+++ N RAL+AAI ASGTDDL N+MFLAFSMS+I+TDL LS Q GWI TITNLGM
Sbjct: 1  MSTLSLDTTNKRALVAAIAASGTDDL NVMFLAFSMTSIMTDLGLSGTQGGWIATITNLGM 60

Query: 61  LVGGLIFGLLADRYNKFKVFKWTILIFSIATGLVFFTTNLSYLYIMRFIAGIGVGGEYGI 120
        LVGGL+FGLLADR++KFKVFKWTIL+FS+ATGL++FT +L YLY+MRFIAGIGVGGEYG+
30  Sbjct: 61  LVGGLLFGLLADRHKKFKVFKWTILIFS VATGLIYFTQSLPYLYLMRFIAGIGVGGEYGV 120

Query: 121 AIAIMAGIVPTNKMGRISSLNGIAGQVGSISSALLAGWLAPALGWRGLFLFGLLPIVLVL 180
        AIAIMAGIVP KMGR+SSLNGIAGQ+GSISSALLAGWLAP+LGWRGLFLFGLLP I+LV+
35  Sbjct: 121 AIAIMAGIVPPEKMGMRSSSLNGIAGQLGSISSALLAGWLAPSLGWRGLFLFGLLPILLVI 180

Query: 181 WMQFAVDDKDILDQYNTDADDEPLDISIKALFDTFVLATQSLALMVMTTVQIAGYFGMMN 240
        WM A+DD+ I D Y + ++ I I L F T L Q+LALMVMTTVQIAGYFGMMN
40  Sbjct: 181 WMTLAIDDQKIWDHYGQEEEECSQPIKINELFKTKSLTAQTLALMVMTTVQIAGYFGMMN 240

Query: 241 WLPTTIQTNLNVSVKNSSLWMIATILGMCLGMLVFGQLLDKFGPRLVYGCFLLSAICVY 300
        WLPTTIQT+LN+SVK+SSLWM+ATI+GMCLGML FGQLLD FG PRL+Y FLL+S+ICVY
45  Sbjct: 241 WLPTTIQTSLNLSVKSSSLWMVATIVGMCLGMLYFGQLLD CFGPRLIYSLFLLASSICVY 300

Query: 301 LFQFATTMPMSMIIGGAVVGFFVNGMFAGYGAMITRLYPHIRSTANNLILNVGRAIGGFS 360
        LFQFA +M SM+IGGA+VGFFVNGMFAGYGAMITRLYPHIRSTANN+ILNVGRA+GGFS
50  Sbjct: 301 LFQFANSMASMVIGGAIVGFFVNGMFAGYGAMITRLYPHIRSTANNVILNVGRALGGFS 360

Query: 361 SVIIGMILDVSNVSMVMLFLASLYIVSFLSMLSILKQLKRQKY 402
        SV IG IL D S +SMVM+FLASLY++SF +M SI QLK ++Y
50  Sbjct: 361 SVAIGSILDASGISMVMI FLASLYVISFGAMWSIGQLKAERY 402

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2033

55 A DNA sequence (GBSx2144) was identified in *S.agalactiae* <SEQ ID 6285> which encodes the amino acid sequence <SEQ ID 6286>. This protein is predicted to be leucyl-tRNA synthetase (leuS). Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

-2291-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3481(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10241> which encodes amino acid sequence <SEQ ID 10242> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00259 GB:AF008220 leucine tRNA synthetase [Bacillus subtilis]
Identities = 569/835 (68%), Positives = 666/835 (79%), Gaps = 42/835 (5%)

Query: 10 YNHKEIEPKWQAFWADNHTFKTGTDASKPKFYALDMFPYPSGAGLHVGHPEGYTATDILS 69
+ HKEIE KWQ +W +N TF T + K KFYALDMFPYPSGAGLHVGHPEGYTATDILS
Sbjct: 3 FQHKIEIEKKWQTYWLENKTFATLDNNEKQKFYALDMFPYPSGAGLHVGHPEGYTATDILS 62

Query: 70 RFKRAQGHNVLHPMGWDAFGLPAEQYAMDGTGNDPAEFTAENIANFKRQINALGFSYDWDR 129
R KR QG++VLHPMGWDAFGLPAEQYA+DTGNDPA FT +NI NF+RQI ALGFSYDWDR
Sbjct: 63 RMKRMQGYDVLHPMGWDAFGLPAEQYALDTGNDPAVFTKQNIIDNFRRIQALGFSYDWDR 122

Query: 130 EVNTTDPNYYKWITQWIFTKLYEKLAYEAEVFNWVEELGTAIANEEVLPDGTSEGGYP 189
E+NTTDP YYKWITQWIF KLYEKLAY EVPVNW LGT +ANEEV+ DG SERGG+P
Sbjct: 123 EINTTDPYYKWITQWIFLKLAYEAEVFNWVEELGTAIANEEVLPDGTSEGGYP 181

Query: 190 VVRKPMRQWMLKITAYAEERLLEDLEEDWPESIKDMQRNWIGKSTGANVTFKVKDTDKDF 249
V R+PM+QWMLKITAYAEERLLEDLEEDWPESIKDMQRNWIG+S GA+V F + D F
Sbjct: 182 VERRPMRQWMLKITAYAEERLLEDLEEDWPESIKDMQRNWIGRSEGAHVHFAIDGHDDSF 241

Query: 250 TVFTTRPDTLFGATYAVLAPEHALVDAITADQAEVAEYKQASLKSDDLARTDLAKEKT 309
TVFTTRPDTLFGATY VLAPEHALV+ ITTA+Q EAV Y ++ KSDL RTDLAK KT
Sbjct: 242 TVFTTRPDTLFGATYTVLAPEHALVENITTAQKEAVEAYIKIQLSKSDDLERTDLAKTKT 301

Query: 310 GVWTGAYAINPVNGKEIPVWIADYVLSYGTGAIMAVPAHDERDWEFAKQFNLDIIPVLE 369
GV+TGAYAINPVNG+++P+WIADYVLSYGTGA+MAVP HDERD+EFK F L + V++
Sbjct: 302 GVFTGAYAINPVNGEKLPIWIADYVLSYGTGA+MAVP HDERD+EFK F L + V++ 361

Query: 370 GGNVEEAAFTEDGLHINSDFLDGLDKAAAIKMWLEAEAGVNEKVITYRLRDWLFSRQR 429
GGNVEEAA+T DG H+NSDFL+GL K AI K++ WLE G +KVITYRLRDWLFSRQR
Sbjct: 362 GGNVEEAAFTEDGLHINSDFLDGLDKAAAIKMWLEAEAGVNEKVITYRLRDWLFSRQR 421

Query: 430 YWGEPIPIHWDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVT-REDGV 488
YWGEPIPIHWDGTSTAVPE ELPL+LP T +I+PSGTGESPLAN+ +W+EVT E G
Sbjct: 422 YWGEPIPIHWDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVT-REDGV 481

Query: 489 KGRRETNTMPQWAGSSWYLYRYPHNTEKLADDELLKQWLPVDIYVGAHAHLHLLYA 548
KGRRETNTMPQWAGS WY+LYRYPHN ++LA E L++WLPVD+Y+GGAHAHLHLLYA
Sbjct: 482 KGRRETNTMPQWAGSCWYFLRYDPHNPDQLASPEKLEKWLVDIYVGAHAHLHLLYA 541

Query: 549 RFWHKVLVDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVETGE 608
RFWHK LYD+GVVPTKEPFQKL+NQGMILG E E
Sbjct: 542 RFWHKFLYDIGVVPTKEPFQKLYNQGMILG-----ENNE 575

Query: 609 ELEQAPAKMSKSLKNVNPDDVVEQYQADTLRVYEMFMGPLDASIAWSEEGLEGSRKFLD 668
KMSKS NVNPD++V +GADTLR+YEMFMGPLDASIAWSE GL+G+R+FLD
Sbjct: 576 -----KMSKSGNVNPDDEIVASHGADTLRVYEMFMGPLDASIAWSEGLDGARRFLD 628

Query: 669 RVYRLI-----TTKEITEENSGALDKVYNETVKAVTEQVDQMKNFNTAIAQLMVFNAA 722
RV+RL +I E L++VY+ETV VT+ + ++FNT I+QLMVFN A
Sbjct: 629 RVWRLFIEDSGELNGKIVEGAGETLERYVHETVMKVTDHYEGLRFNTGISQLMVFINEAY 688

Query: 723 KEDKLFSDYAKGFVQLIAPFAPHLGEELWQVLTASGQSISYVPWPSYDESKLVENEIEIV 782
K +L +Y +GFV+L++P APHL EELW+ L SG +I+Y WP YDE+KLV++E+EIV
Sbjct: 689 KATELPKEYMEGFVKLLSPVAPHLAEELWEKLGHS+TIAYEAWPVYDETKLVDEVEIV 747

Query: 783 VQIKGKVKAKLVVAKDLSREELQDLALANEKVQAEIAGKDIKVIAPNKLNVIV 837
VQ+ GKVKAKL V D ++E+L+ LA A+EKV+ ++ GK I K+IAPV KLVNIV

-2292-

Sbjct: 748 VQLNGKVKAKLQVPADATKEQLEQLAQADEKVKEQLEGKTIRKIIIVPGKLVNIV 802

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6287> which encodes the amino acid sequence <SEQ ID 6288>. Analysis of this protein sequence reveals the following:

5 Possible site: 46
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.4358(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 815/833 (97%), Positives = 827/833 (98%)

15 Query: 7 MTFYNHKEIEPKWQAFWADNHTFKTGT DASKPKFYALDMFPYPGAGLHVGHPEGYTATD 66
 MTFY+H IEPKWQAFWADNHTFKTGT DASKPKFYALDMFPYPGAGLHVGHPEGYTATD
 Sbjct: 1 MTFYDHTAIEPKWQAFWADNHTFKTGT DASKPKFYALDMFPYPGAGLHVGHPEGYTATD 60

20 Query: 67 ILSRFKRAQGHNVLHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD 126
 ILSRFKRAQGHNV+LHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD
 Sbjct: 61 ILSRFKRAQGHNVILHPMGWDAFGLPAEQYAMDTGNDPAEFTAENIANFKRQINALGFSYD 120

25 Query: 127 WDREVNITDPNYYKWTQWIFTKLYEKGLAYEAEVFPVNWVEELGTAIANEEVLPDGT SERG 186
 WDREVNITDPNYYKWTQWIFTKLYEKGLAYEAEVFPVNWVEELGTAIANEEVLPDGT SERG
 Sbjct: 121 WDREVNITDPNYYKWTQWIFTKLYEKGLAYEAEVFPVNWVEELGTAIANEEVLPDGT SERG 180

30 Query: 187 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTD 246
 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTD
 Sbjct: 181 GYPVVRKPMRQWMLKITAYAERLLEDLEEVDWPESIKDMQRNWIGKSTGANVTFKVKDTD 240

35 Query: 247 KDFTVFTTRPDTLFGATYAVLAPEHALVDAITADQAEAVA EYKRQASLKS DLARTDLAK 306
 KDFTVFTTRPDTLFGATYAVLAPEHALVDAITADQAEAVA+YKRQASLKS DLARTDLAK
 Sbjct: 241 KDFTVFTTRPDTLFGATYAVLAPEHALVDAITADQAEAVAKYKRQASLKS DLARTDLAK 300

40 Query: 307 EKTGVVWTGAYAINPVNGKEIPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFNLDIIP 366
 EKTGVVWTGAYAINPVNG E+PVWIADYVLASYGTGAIMAVPAHDERDWEFAKQF LDIIP
 Sbjct: 301 EKTGVVWTGAYAINPVNGNEMPVWIADYVLASYGTGAIMAVPAHDERDWEFAKQFKLDIIP 360

45 Query: 367 VLEGGNVEEAFTEDGLHINSDFLDGLDKAAAIKMWLEAEAGVGNEKVITYRLRDWLFS 426
 VLEGGNVEEAFTEDGLHINS FLDGLDKA+AIKMWLEAEAGVGNEKVITYRLRDWLFS
 Sbjct: 361 VLEGGNVEEAFTEDGLHINSGLDGLDKASAIKMWLEAEAGVGNEKVITYRLRDWLFS 420

50 Query: 427 RQRYWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANLTDWLEVTRED 486
 RQRYWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLAN+TDWLEVTRED
 Sbjct: 421 RQRYWGEPIPIIHWEDGTSTAVPESELPLVLPVTKDIRPSGTGESPLANVTDWLEVTRED 480

55 Query: 487 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLAD EELLKQWLPVDIYVGGA EHAVLHLL 546
 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLAD EELLKQWLPVDIYVGGA EHAVLHLL
 Sbjct: 481 GVKGRRETNTMPQWAGSSWYYLRYIDPHNTEKLAD EELLKQWLPVDIYVGGA EHAVLHLL 540

60 Query: 547 YARFVHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET 606
 YARFVHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET
 Sbjct: 541 YARFVHKVLYDLGVVPTKEPFQKLFNQGMILGTSYRDSRGALVATDKVEKRDGSFFHVET 600

65 Query: 607 GEELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRK 666
 GEELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRK
 Sbjct: 601 GEELEQAPAKMSKSLKNVNPDDVVEQYGADTLRVYEMFMGPLDASIAWSEEGLEGSRK 660

Query: 667 LDRVYRLITTTKEITEENSGALDKVYNETVKAVTEQVDQMKNFTAI AQLMVFN AANKEDK 726
 LDRVYRLITTTKEITEENSGALDKVYNETVKAVTEQVDQMKNFTAI AQLMVFN AANKEDK
 Sbjct: 661 LDRVYRLITTTKEITEENSGALDKVYNETVKAVTEQVDQMKNFTAI AQLMVFN AANKEDK 720

Query: 727 LFS DYAKGFVQLIAPFAPHLGEELWQVLTASQSISYVPWPSYDESKLVENEIEIVVQIK 786
 LFS DYAKGFVQLIAPFAPHLGEELWQ LTASG+SISYVPWPSYDESKLVEN++EIVVQIK

-2293-

Sbjct: 721 LFSDYAKGFVQLIAPFAPHLGEELWQALTASGESISYVPWPSYDESKLVENDVEIVVQIK 780

Query: 787 GKVKAKLVVAKDLSREELQDLALANEKVQAEIAGKDIKVIAPNKLNVNIVK 839
 GKVKAKLVVAKDLSREELQ++ALANEKVQAEIAGKDIKVIAPNKLNVNIV+K

5 Sbjct: 781 GKVKAKLVVAKDLSREELQEVALANEKVQAEIAGKDIKVIAPNKLNVNIVK 833

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2034

10 A DNA sequence (GBSx2145) was identified in *S.agalactiae* <SEQ ID 6289> which encodes the amino acid sequence <SEQ ID 6290>. This protein is predicted to be KIAA1074 protein. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

15

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8961> which encodes amino acid sequence <SEQ ID 8962> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7

SRCFLG: 0

25

McG: Length of UR: 19

Peak Value of UR: 2.86

Net Charge of CR: 4

McG: Discrim Score: 10.27

GvH: Signal Score (-7.5): -3.61

30

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

Amino Acid Composition: calculated from 1

ALOM program count: 0 value: 2.12 threshold: 0.0

PERIPHERAL Likelihood = 2.12 7

35

modified ALOM score: -0.92

*** Reasoning Step: 3

----- Final Results -----

40

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8962 (GBS117) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 8; MW 22.5kDa).

GBS117-His was purified as shown in Figure 200, lane 7.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50

-2294-

Example 2035

A DNA sequence (GBSx2146) was identified in *S.agalactiae* <SEQ ID 6291> which encodes the amino acid sequence <SEQ ID 6292>. This protein is predicted to be YirC (resE). Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -10.88    Transmembrane 177 - 193 ( 173 - 196)
      INTEGRAL    Likelihood = -4.09     Transmembrane 10 - 26 ( 5 - 29)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15  The protein has homology with the following sequences in the GENPEPT database.
    >GP:CAB15292 GB:Z99120 similar to two-component sensor histidine
      kinase [YvqA] [Bacillus subtilis]
      Identities = 108/379 (28%), Positives = 193/379 (50%), Gaps = 33/379 (8%)

20  Query: 92  DNHHKESHDIIRYLTKRLWQISKEKDMFVTIKKTTYVMTKDYSGILVDGSIKKVPKA 151
      +N + S + L+ + ++ K D KKK Y + D +G V IKK
      Sbjct: 86  ENEEASSDKDLSILSSSFHVKVYLADKQ--EAKKKRY---SADVNGEKVFFVIKKGLSV 140

25  Query: 152 QSQLFHVINF-----DITYTQHLITKINHFLIVLVLTYPMLFIMRKTFTGIRESIQ 205
      Q +++++ D+ YT L ++ + V+++L++IP +++ + + +
      Sbjct: 141 NGQSAMMLSYALDSYRDDLAYT--LFKQLLFITIAVVILLSWIPAIWLAKY----LSRPLV 194

30  Query: 206 SVQTYISSLWKNQGNHQSSQKEIVFSDFDPLLESQEMANRIYQAEESQRNFFQNASHL 265
      S + ++ + ++ + K + L +EM ++ Q +E++R QN SH+L
      Sbjct: 195 SFEKHVKRI--SEQDWDVPVKVDRKDEIGKLGHTIEEMRQKLVQKDETERILLQNISHDL 252

35  Query: 266 RTPLMISIQGYTEGVQEGII---DAELAHSVILQESKKMKQLVDDIILLSKLD--SNLSDQ 320
      +TP+M I+GYT+ +++GI D E VI E+ K+++ + D++ L+KLD + Q
      Sbjct: 253 KTPVMVIRGYTQSIKDGIFPKGDLNFTVDVIECEALKLEKKIKDLLYLTKLDYLAQKQKVQ 312

40  Query: 321 KDEFSLNELNLSIIAYFKPLANKQKISITYRDPKHEKLLK-GNEELIQRAINNILSNALR 379
      D FS+ E+ +I K A K+ +++ D E +L G+ E + + NIL N +R
      Sbjct: 313 HDMFSIVEVTEEVIERLK-WARKE---LSWEIDVEEDILMPGDPEQWNKLENIENQIR 368

45  Query: 380 YAVSHIEISYT----NQKLTISNDGPAISKEDLPYIFDRFYKGHGGQTGIGLAMTKKEIK 435
      YA + IEIS N +TI NDGP I E L +++ F KG G+ GIGL++ K I+
      Sbjct: 369 YAEKIEISMKQDDRNIVITIKNDGPHIEDEMLSSLYEPFNKGKGEFGIGLSIVKRILT 428

50  Query: 436 QHHGNIIAESDSTSTTFTI 454
      H +I E+D T ++ I
      Sbjct: 429 LHKASISINDKTGVSYRI 447

```

There is also homology to SEQ ID 1178.

SEQ ID 6292 (GBS279) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 7; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 6; MW 79.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2295-

Example 2036

A DNA sequence (GBSx2147) was identified in *S.agalactiae* <SEQ ID 6293> which encodes the amino acid sequence <SEQ ID 6294>. This protein is predicted to be two-component response regulator (mtrA). Analysis of this protein sequence reveals the following:

5 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1706(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10239> which encodes amino acid sequence <SEQ ID 10240> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05663 GB:AP001513 two-component response regulator [Bacillus halodurans]
 Identities = 87/220 (39%), Positives = 124/220 (55%), Gaps = 4/220 (1%)

20 Query: 11 IYFADDEKNIRDLVVPFLEHGDFTVRAFETGDLLEAYKNQKPDVLILDIMPGTNGLDV 70
 I DDE ++R+LV +L +GF V ETGD ++ + + DLV+LD+MM +G
 Sbjct: 7 ILIVDDELDLRELVTSYLRKEGFAVYTAETGDEAIKRLEQEPMDLVVLDVMMDEMDGFTA 66

25 Query: 71 MKSIRQYDNIPIIMLTARSDVDFTAFNLGTDDYFTKPFSPIKLSLHVKALFKRLDEKA 130
 K IR + IPIIMLTAR + D + +G DDY KPFSP +L ++ +R
 Sbjct: 67 CKEIRAFSQIPIIMLTARGGEDDKVMGLQIGADDYIVKPFSPRELVARIEVALRRTQGIQ 126

30 Query: 131 IKNDTQYQFLDLTLDTEKRIALLSNEEMPLTKTEFDLLVLEKPEAFSRETLLNRIWG 190
 +DT Y+F +L + R ++ +E+ LTK E+D L+ L+E F+RE L +R+WG
 Sbjct: 127 QVDDTGYRFNELRIQPSGRKVFVNGQEISLTKKEYDLLVFLEHGRVFTREHLHDLRWG 186

 Query: 191 FDDIES--RAVDDTIKRLRKKFKQYHSQVSIKTVWGYGFK 228
 D + R VD IK LR K K + IKTWVG G+K
 Sbjct: 187 MDTQQGTLRTVDTHIKTLRLKLP--ADRFIKTVWGVGYK 224

35 There is also homology to SEQ ID 3260.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2037

40 A DNA sequence (GBSx2148) was identified in *S.agalactiae* <SEQ ID 6295> which encodes the amino acid sequence <SEQ ID 6296>. Analysis of this protein sequence reveals the following:

 Possible site: 55
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.18 Transmembrane 1568 -1584 (1568 -1585)
 INTEGRAL Likelihood = -0.16 Transmembrane 338 - 354 (338 - 354)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 10237> which encodes amino acid sequence <SEQ ID 10238> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2296-

>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
Identities = 797/1594 (50%), Positives = 1056/1594 (66%), Gaps = 39/1594 (2%)

5 Query: 21 MNTKQRFSSIRKYKLGAVSVLLGTLFFLGGITNVAADSVINKPSDIAVEQQVKDSPTS-IA 79
M K+ FS+RKYK+G VSVLLG +F G +VAAD + + + VE V D+ S A
Sbjct: 1 MKKKEFTSLRKYKIGTVSVLLGAVFLFAGAPSVAADELTSLV-ETKVEATVPDAIVSESA 59

10 Query: 80 NETPTNN--TSSALASTAQDNLVTKANNSPTETQPVAESHQATETTFSPVANQPVESTQE 137
+E+P +++ +T+ D T ++ + S + ET P P S ++
Sbjct: 60 SESPVVEELVDTSVSATSTDVTTTNEETPGSEALENSANTEVETTQPAVETPAISEKK 119

15 Query: 138 VSKTPLTKQNLA VKSTPAISKET--PQNIDSNKIITVPKVVNTGYKGE GTVVVAIIDSGLD 195
V + K ++A ++T ++E PQNIDSN IITVPKVV +GYKGE GTVVVAIIDSGLD
Sbjct: 120 VEEEE--KLSVADETTAITNQEEAKPQNIDSNKIITVPKVVWISGYKGE GTVVVAIIDSGLD 177

20 Query: 196 INHDALQLNDSTKAKYQNEQQMNAKAKAGINYGKWNKVI FGHNYVDVNTLKEVKST 255
++HD L ++D + AKY++E+++ AAK AGI YG+W+N+KV+FG+NYVDVNT LKE
Sbjct: 178 VDHDVLHISDLSTAKYKSEKEIEAAKEAAGITYGEWFNDKVVFYNYVDVNTVLKEEDKR 237

25 Query: 256 SHGMHVTSIATANPSKKDTNELIYGVAPEAQVMFMRVFSDEKRGTPALYVKAIEDAVKL 315
SHGMHVTSIAT NP++ +L+YGAPEAQVMFMRVFS D K TG ALYVKAIEDAVKL
Sbjct: 238 SHGMHVTSIATGNPTQPVAGQLMYGVAPEAQVMFMRVFS DLKATTGAALYVKAIEDAVKL 297

30 Query: 316 GADSINLSLGGANGSLVNADDRLIKALEMARLAGSVVIAAGNDGTFGSGASKPSALYPD 375
GADSINLSLGGANGS+VN ++ + A+E AR AGSVVIAAGNDGTFGSG S PSA YPD
Sbjct: 298 GADSINLSLGGANGSVVMNENVTAAIEAARRAGSVVIAAGNDGTFGSGHSNPSADYPD 357

35 Query: 376 YGLVGSPTAREAISVASYNNTTLVNKVFNIIGLENNRNLNGLAAYADPKVSDKTFEFG 435
YGLVG+PSTA +AISVASYNNT+ +KV NIIGLENN +LN G +++ +P+ S FE+G
Sbjct: 358 YGLVGAPSTAHDAISVASYNNTTVGSKVINIIGLENNADLNKSSFDNPEKSPVPFEIG 417

40 Query: 436 KQYDYVVFVGKNDNDYKDKTLNGKIALIERGDITFTKKVMAINHGAVGAIIFNNKAGEA 495
K+Y+YV+ G G +D+ L GK+ALI+RG ITF++K+ NA GAVG +IFN++ GEA
Sbjct: 418 KEYEYVYAGIQASDFDGLDLTGKALIKRGITITFSEKIANATAAGAVGVVIFNSRPGEA 477

45 Query: 496 NLTMSLDPEASAIPAIFTQKEFGDVLAKNNYKIVFNNIKNKQANPNAGVLSDFSSWGLTA 555
N++M LD A AIP++F EFG+ LA N+YKI FNN + + NP AG+LSDFSSWGL+A
Sbjct: 478 NVSMQLDDTAIAIPSVFIPLEFGEALAANSYKIAFNNETDIRPNEAGLLSDFSSWGLSA 537

50 Query: 556 DGQLKPDLSAPGGSIIYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPELKKGDI 615
DG+LKPDL+APGG+IYAAINDN+Y M GTSMASPHVAGA LVKQYLL +P +I
Sbjct: 538 DGELKPDLAAPGGAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTKSPQEI 597

55 Query: 616 ERTVKYLLMSTAKAHLNKDTGAYTSRQQGAGIIDVAAAVQTGLYLTGGENNYGSVTLGN 675
E VK+LLMSTAKAH+NK+T AYTSPRQQGAGIID AAA+ TGLYLT GE+ YGS+TLGN
Sbjct: 598 EALVKHLLMSTAKAHVNKETTAYTSRQQGAGIIDTAAISTGLYLT-GEDGYGSITLGN 656

60 Query: 676 IKDKISFDVTVHNINKVAKDLHYTTYLTNDQVKDGFVTLAPQQLGTFTGKTIRIEPGQTQ 735
++D SF VT+HNI K L+Y+T L TD + L + + + + + +
Sbjct: 657 VEDTFSFTVTLHNITNEDKTLNYSQTLTDTAQKRIDLHLSGTSISRDSWRKVTVKANSS 716

65 Query: 736 TITIDIDVSKYHMLKKVMPNGYFLEGYVRFTDPVDGGEVLSIPYVGFGEFQNLLEVLEK 795
T+TI++D S + + L +M NGY+LEG+VRFTD D G+++SIPYVGF+GEFQNL VLE+
Sbjct: 717 TVTINVDASSF AEELTGLMKNGYILEGFVRFTDVADDGDIVSIPYVGFGEFQNLAVLEE 776

70 Query: 796 SIYKLVANKEKGFYFQP--KQINEVPGSEDTALMTTSSEPIYSTDGTSPQLKALGSYK 853
IY L+A+ + GFYF+P Q N V S YT L+T S+E IYSTD S +K LG++K
Sbjct: 777 PIYNLIADGKGGFYFEPVTAQNTVDISHHYTGLVTGSTELIYSTDKRSDSAIKTLGTFK 836

75 Query: 854 SIDGKWILQLDQKGPHLAISFNDDQNQDAVAVKGVFLRNFNRLRAKVYRADDVNLOKPL 913
+ G ++L+LD+ G+PHLAISEN D NQD++ KGVFLRN+ +L A VY ADD PL
Sbjct: 837 NKAGYFVLELDESCKPHLAISPNGDDNQDSLFGVFLRN+YDLVASVYAADDERTNPL 896

80 Query: 914 WVSAPQAGDKNYSGNTENPKSTFLYDTEWKGTTTDGIPLDYGKYVLTYYSDVPGSKP 973
W S PQ+GDKN YSGN +NPKS+ +Y TEW GT +DG L DGKY+YVLT Y S VPG+
Sbjct: 897 WESQPQSGDKNIYSGNPKNPKSSIIYPTWNGTSDGNALADGKYQYVLTYYSSKVPGA 956

85 Query: 974 QQMVFDITLDRQAPTTLTATYDKDRIFKARPAVEHGESGIFREQVYLLKKDKDGHNSV 1033

-2297-

Q M+FD+ +DR++P +TTATYD+ F RPA+E GESG++REQVFYL D G ++
 5 SbJct: 957 QTMIFDVIIDRESPVITTTATYDETNFTFNERPAIEKGESGLYREQVFYLVDASG-VTTI 1015

Query: 1034 LRQQGEDGILVEDNKVFIKQEKDGSFILPKEVNDFSHVYYTVEDYAGNLVSAKLEDLINI 1093
 + V DNKVF+ Q DGSF LP ++ D S YYTVEDYAGN+ K+E+LI+I
 10 SbJct: 1016 PSLKNGDVTVDNKNVFAQNDDGSFTLPLDLADISKFYTVEDYAGNISYEKVENLISI 1075

Query: 1094 GNKNGLVNVKVFSPELNSNVDIDFSYSVKDDKGNIIKK-QHHGKDLNLLKLPFGTYTDFL 1152
 GN+ GLV V + + NS V I FSYSV D+ G I+ + + D ++LKLPGTYTDFL
 10 SbJct: 1076 GNEKGLVTVNILDKDNTSPVPILFSYSVTDETCKIVAELEPRYAGDTSVLKLPFGTYTDFL 1135

Query: 1153 FLYDEERANLISPKSVTVTISEKDSLKDVLFKVNLKKAALLVEFDKLLPKGATVQLVTK 1212
 FLYD E ++L VTI E +S +V F V L KA LL++ D LLP G+T+QLVT
 15 SbJct: 1136 FLYDTEWSSLAGEKAVVITLEDNSTAEVNFYVTLKDKANLLIDIDALLPSGSTIQLVTA 1195

Query: 1213 TNTVVDLPKATYSPTDYGNIPVGDYRLNVTLPSTYENLDDLLSVKEDQVNLTKLT 1272
 + LP A YS TDYK +PVG Y + TLP GY LE LD V+V +Q N+ KLT
 20 SbJct: 1196 DGQAIQLPNKYSKTDYKGFVPVGTYYTILPTLPEGYEFLEELD---VAVLANQSNVKKLT 1252

Query: 1273 LINKAPLINALAEQTDITQPVFYNAGTHLKNNYLANLEKAQTLIKNRVEQTSIDNAIAA 1332
 LINK L +AE + +YNA L+ Y LE A + N+ Q +D+A+A+
 25 SbJct: 1253 LINKVALKELIAELAGLEETARYYNASPELQTAYAKALEDANAVYANKHNQAQVDSALAS 1312

Query: 1333 LRESRQALNGKETDTSLLAKAILAETEIKGNYQFVNASPLSQSTYINQVQLAKNLLQKPN 1392
 L +R+ LNG+ TD L + T + N+ + NA Q Y V+ A+ +L + N
 25 SbJct: 1313 LVAAREQLNGQATDKEKILAEVSNYPTPTQANFIYNAENTKQIAYDTAVRSAQLVLNQEN 1372

Query: 1393 VTQSEVDKALENLDIAKNQLNGHETDYSGLHMIIKANVLKQTSSKYQNASQFAKENYNN 1452
 VTQ+ V++AL +L AK L+G +TD S L + ++VLK T +KY NAS+ K+ Y+
 30 SbJct: 1373 VTQAVVNQALADLLAAKANLDGQKTDISALRSVSVSVLKATDAKYLNASENVKQAYDQ 1432

Query: 1453 LIKKAELLLSNRQATQAQVEELLNQIKATEQEELDG----RDRVSSAENYSQSLNDNDSLN 1508
 ++ A+ +L + A+QA V++ L + + + ELDG + N + D ++
 35 SbJct: 1433 AVEAAKAILVDESASQASVDQALAVLTSQAQELDG VATSTNDAKEPANTATDKKDEGTVT 1492

Query: 1509 TTPIN-----PP-----NQPQALIFKKGMTKESEVAQKRVLGVTSTQTDNQVKTKNL 1555
 PI+ PP N I +K + + + L + + NQ+ + +L
 35 SbJct: 1493 PPPIDSEIVDVQAPPVKDTGNSEHVPIGQK-PNPQPTLPRPVTQLASLSSPNQEKQVTL 1551

Query: 1556 PKTGESTPKITYTILLFSLMLGLATIKLSIKR 1589
 P TGE+ K L ++GL T+ L SI+R
 40 SbJct: 1552 PNTGENDTK----YYLVPGVIIIGLGT-LVSIRR 1580

A related GBS gene <SEQ ID 8963> and protein <SEQ ID 8964> were also identified. Analysis of this
 45 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 SRCFLG: 0
 50 McG: Length of UR: 1
 Peak Value of UR: 2.55
 Net Charge of CR: 4
 McG: Discrim Score: 2.60
 GvH: Signal Score (-7.5): -0.78
 Possible site: 35
 >>> Seems to have a cleavable N-term signal seq.
 55 Amino Acid Composition: calculated from 36
 ALOM program count: 1 value: -0.16 threshold: 0.0
 INTEGRAL Likelihood = -0.16 Transmembrane 318 - 334 (318 - 334)
 PERIPHERAL Likelihood = 2.54 1161
 modified ALOM score: 0.53
 60 icml HYPID: 7 CFP: 0.106

*** Reasoning Step: 3

----- Final Results -----
 65 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

Streptococcu

GP|9963932| cell envelope proteinase Insert characterized

GP|9963932|gb|AAG09771.1|AF243528_1|AF243528(1 - 1584 of 1585) cell envelope proteinase {*Streptococcus thermophilus*}

%Identity = 50.4 %Similarity = 67.4

Matches = 794 Mismatches = 498 Conservative Sub.s = 267

255 285 315 345 375 405 435 465
KNALGTVLNLQPQNNL**KFRKL*KILIFYVLIVFVLIIMLQEKEIFMNTKQRFSSIRKYKLGAVSVLLGLTFFLGGITNVAA
| : ||:|||:| ||||| ::: | :||
MKKKETFSLRKYKIGTVSVLLGAVFLFAGAPSVAA
10 20 30

495 525 552 576 606 636 666 696
DSVINKPSDIAVEQQVKDSPTS-IANETPT--NTSSALASTAODNLVTKANNSTPTETQFVAESHSQATETTFSPVANQPV
| : : || | : | |:| : :: : |: | | : : : | : || | |
DE-LTSLVETKVEATVPDAIVSESASESPVVEELVDTSEATSTDVTTTNDNEETPGSEALENSANTEVETTQPAVETPA
50 60 70 80 90 100 110

```

726          756          780          810          840          870          900          930
ESTQEVSKTPLTKQNLA VKSTPAISKE--TPQNIDSNKIITVPKWVNTGYKKGEGTVVAIIDSGLDINHDA LQ LNDSTKAK
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
ISEKKVEEEE--KLSVADETTAITNQEEAKPQNIDSNITVPKWVSYGKKGEGTVVAIIDSGLDVHDVLHISDLSTAK

```

960 990 1020 1050 1080 1110 1140 1170
YQNEQQMNAAKAGINYGKWNKKVIFGHNYVDVNTLKEVKSTSHGMHVTSIATAPNSKKDTNELIYGVAPEAQVMFM
|:::|| | | | |:|::||:||||| || | ||||| | :| :|:|||||
YKSEKEIEAAKEAAGITIGEWFNDKVVFGYNYVDVNTVLKEEDKRSHGMHVTSIATGNPTQPVGALMLGYVAPEAQVMFM

 210 220 230 240 250 260 270

1200 1230 1260 1290 1320 1350 1380 1410
RVFSDEKRGTPGALYVKAIEDAVKLGADSINLSLGGANGSLVNADDRLIKALEMARLAGVSVVIAAGNDGTFGSGASKPS
||||| | || ||||| ||||| ||||| :|| :: : |:| ||||| |||||
RVFSDLKATGAALYVKAIEDAVKLGADSINLSLGGANGSVVMNNVENTAAIEARRAGVSVVIAAGNDGTFGSGHSNPS
 290 300 310 320 330 340 350

1440 1470 1500 1530 1560 1590 1620 1650
ALYPDYGLVGSPPSTAREAISVASYNNTTLVNKVFNIIGLENNRNLNNGLAAYADPKVSDKTFEYVGKQYDYVFVGKGNND
| | | | | : | | | : | | | | | : | | | | | : | | : : : | | : | : | : | : | : | : |
ADYPDYGLVGAPSTAHDIAISVASYNNTTVGSKVINIIGLENNADLNYGKSSFDNPEKSPVPFTEIGKEYEYVYAGIGQASD
370 380 390 400 410 420 430

1680 1710 1740 1770 1800 1830 1860 1890
YKDKTLNGKIALIERGDIITFTKKVVNAINHGAVGAIIFNNKAGEANLTMSLDPEASAIPAIFTQKEFGDVLAKNNYKIVF
:
FDGLDLTGKIALIKRGTITTFSEKIANATAAGAVGVVIFNSRPGEANVMQLDDTAIAIPSVFIPLFEGEALAAANSYKIAT

450 460 470 480 490 500 510

1920 1950 1980 2010 2040 2070 2100 2130
NNIKNKQANPNAGVLSDFSSWGLTADGQLKPDLSAPGGSIYAAINDNEYDMMSGTSMASPHVAGATALVKQYLLKEHPEL
|| : : || |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
NNETDIRPNPEAGLLSDFSSWGLSADGELKPDLAAPGGATYAAINDNDYANMQGTSMASPHVAGA AVL VKQYLLAT YPTK

530 540 550 560 570 580 590

2160 2190 2220 2250 2280 2310 2340 2370
 KKGDIERTVKYLMLSTAKAHLNKDTGAYTSPROOGAGIIDVAAVOTGLYLTTGGENNYGSVTLGNIKDKISFDVTVHNIN

-2300-

```

      1480      1490      1500      1510      1520      1530      1540
4980      5010      5040      5070      5100      5130      5160      5190
KTNKLPKTGESTPKITYTILLFSLSMLGLATIKLSIKRE*NTLKNRRARHQLLAINS**LVPF*GA*NDVPKDLFSVSW
5  : :|| |||: | |:: : || : :: | |
QVTQLPNTGENDTK--YYLVPGVIIGLGTLLVSIIRRHKEEV
      1560      1570      1580

```

A related GBS nucleic acid sequence <SEQ ID 10965> which encodes amino acid sequence <SEQ ID 10966> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6297> which encodes the amino acid sequence <SEQ ID 6298>. Analysis of this protein sequence reveals the following:

LPXTG motif: 1614-1619

Possible site: 33

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -4.46 Transmembrane 1623 -1639 (1621 -1641)

----- Final Results -----

```

bacterial membrane --- Certainty=0.2784(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

>GP:AAG09771 GB:AF243528 cell envelope proteinase [Streptococcus thermophilus]
Identities = 465/1125 (41%), Positives = 668/1125 (59%), Gaps = 61/1125 (5%)

```

Query: 1  VEKKQRFSLRKYKSGTFSVLIGSVFLVM-TTTVAADELSTMSEPTITNHAQQQAQHLTNT 59
++KK+ FSLRKYK GT SVL+G+VFL +VAADEL+++ E + T
Sbjct: 1  MKKKETFSLRKYKIGTVSVLLGAVFLFAGAPSVAADELTSLVETKVEA-----T 49

```

```

Query: 60  ELSSAESKSQDTSQITLKTNREKESQDLVSEPTTTELADTDAASMANTGSDATQKSASL 119
+ S+S S + E+ D E T+T++ TD GS+A + SA
Sbjct: 50  VPDAIVSESASESPVV-----EELVDTSVEATSTDVTTTNDNEE-ETPGSEALENSA-- 99

```

```

Query: 120 PPVNTDVHDWVKTKGAWDKGYKGQKQVVAVIDTGIDPAHQSMRISDVSTAKVSKEDMLA 179
NT+V T+ A + + KV + + ++D +TA +E
Sbjct: 100 ---NTEVET---TQPAVETPAISEKKV-----EEEEKLSVADETTAITNQEE---- 140

```

```

Query: 180 RQKAAGINYSWINDKVFVFAHNYVENSNDNIKE-NQFEDFDEDWENFEFDAEAPKAIKKH 238
K I+ + I V+ Y + + D D D + + A+ K+ K+
Sbjct: 141 -AKPQNIDSNTIITVPKWYSGYKGEQTVVAIIDSGLDVDHDLHISDLSTAKYKSEKEI 199

```

```

Query: 239 KIYRPQSTQAPKETVIKTEETDGSHDIDWTQTDDDTKYESHGMHVTGIVAGNSKEAAATG 298
+ + + E + G + +D + SHGMHVT I GN + A G
Sbjct: 200 EAAKEAAGITYGEW-FNDKVVFQYNYVDVNTVLKEEDKRSHGMMHVTSIATGNPTQPAV-G 257

```

```

Query: 299 ERFLGIAPEAQVMFMRVVFANDIMGSAESLFIKAIEDAVALGADVINSLSGTANGAQLSGS 358
+ G+APEAQVMFMRV++ + +L++KAIEDAV LGAD INLSLG ANG+ ++ +
Sbjct: 258 QLMYGVAPAEQVMFMRVFSDLKATTGAALYVKAIEDAVKLGADSINLSLGGANGSVVMNM 317

```

```

Query: 359 KPLMEAIEKAKKAGVSVVVAAGNERVYSGSDHDDPLATNPDYGLVGSPTGRTPTSVAAIN 418
+ + AIE A++AGVSVV+AAGN+ +GS H +P A PDYGLVG+PST SVA+ N
Sbjct: 318 ENVTAIEAARRAGVSVVIAAGNDGTFGSGHSNPSADYPDYGLVGAPSTAHDASVSYN 377

```

```

Query: 419 SKWVIQRLMTVKELENRADLNHGKAIYSESVDKDIKDSLGYDKSHQFAYVKESTDAGYN 478
+ V +++ + LEN ADLN+GK+ + ++ + + +G + + +A + +++D ++
Sbjct: 378 NTTVGSKVINIIGLENNADLNNGKSSF-DNPEKSPVPFEIGKEYEYVYAGIGQASD--FD 434

```

```

Query: 479 AQDVKGKIALIERDPNKTYDEMIALAKKHGALGVLIFFNNKPGQSNRSMRLTANGMGIPSA 538
D+ GK+ALI+R T+ E IA A GA+GV+IFN++PG++N SM+L + IPS
Sbjct: 435 GLDLTGKLALIKRG-TITFSEKIANATAAGAVGVVIFNSRPGEANVSMQLDDTAIAIPSV 493

```

-2301-

5 Query: 539 FISHEFGKAMSQLNGNGTGSLEFDSVSKAPSQKGNEMNHFSNWLTS DGYLKP DITAPG 598
FI EFG+A++ + + F++ P+ + ++ FS+WGL++DG LKPD+ APG
Sbjct: 494 FIPLEFGEALAA---NSYKIAFNNETDIRNP EAGLLSDFSSWGLSADGELKPD LAAPG 549

10 Query: 599 GDIYSTYNDNHYSQTGTSMAS PQIAGASLLVKQYLEKTQPNLPKEKIADIVKNLLMSNA 658
G IY+ NDN Y + GTSMASP +AGA++LVKQYL T P ++I +VK+LLMS A
Sbjct: 550 GAIYAAINDNDYANMQGTSMASPHVAGAAVLVKQYLLATYPTKSPQEIEALVKHLLMSTA 609

15 Query: 659 QIHVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTGKDNYSISLGNITD TMTFDVTVHN 718
+ HVN ET TSPRQQGAG+++ A+++GLY+TG+D YGSI+LGN+ DT +F VT+HN
Sbjct: 610 KAHVNETTAYTSPRQQGAGI IDTAAAI STGLYLTGEDGYGSI TLGNVEDTFSFTVT LHN 669

20 Query: 719 LSNKDKTLRYDTELLTDHVDPQKGRFTLTSHSLKTYQGGEVTPANGKVTVRV TMDVSQF 778
++N+DKTL Y T+L TD + TS S +++ +VTV AN TV + +D S F
Sbjct: 670 ITNEDKTLNYSQT LTDTAQKRIDLHGSTSISRDSWR--KVTVKANSSTT VTINVDASSF 727

25 Query: 779 TKELTKQMPNGYYLEGFVFRFDSQDDQLNRVNI PFVGFKGQFENLAVAEESYRLKSQ GK 838
+ELT M NGYYLEGFVRF D DD + V+IP+VGF+G+F+NLA V EE IY L + GK
Sbjct: 728 AEELTGLMKNGYYLEGFVRFDTVADDG-DIVSIPYVGFGEFQNLAVLEEPIYNLIADGK 786

30 Query: 839 TGFYFDE-SGPKDDIYVGKHTGLVTLGSETNVSTKTISDNGLHTLGT FKNADGKFILEK 897
GFYF+ + + + + H+TGLVT +E ST SD+ + TLGTFKN G F+LE
Sbjct: 787 GGFYFEPVTAQPN TVDISHHYTGLVTGSTELIYSTDKRSDSAIKTLGTFKNKAGYFVLEL 846

35 Query: 898 NAQGNPVLAI SPNGDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVS-PESFKG 956
+ G P LAISPNGD+NQD FKGVFLR Y L ASVY A D E NPLW S P+S G
Sbjct: 847 DESGKPHLAISPNGDDNQDSL VFKGVFLRNYTDLVASVYAADDTERTNPLWESQPQS--G 904

40 Query: 957 DKN-FNSDIRFAKSTTLLGTAFSGKSLTGAELPDGHYHYVVSYYPDVVGAKRQEMTFDMI 1015
DKN ++ + + KS+ + T ++G G L DG Y YV++Y V GA Q M FD+I
Sbjct: 905 DKNISGNPNKPNKSSIIYPTWNGTDS DGNALADGKYQYVLTYSKVPGA AVQTMIFDVI 964

Query: 1016 LDRQKPVLSQATFD PETNRFKPEPLKDRGLAGVRKDSVFYLERKDNKPYT VTINDSYKYV 1075
+DR+ PV++ AT+D F P P ++G +G+ ++ VFYL + T+ V
Sbjct: 965 IDRESPVITTATYDETNTFTNPRPAIEKGESGLYREQVFYLVADASGVTTIPSL LKNGDV 1024

Query: 1076 SVEDNKTFVERQADGSFILPLDKAKLGDFYVMVEDFAGNVAIAKL 1120
+V DNK FV + DGSF LPLD A + FYY VED+AGN++ K+
Sbjct: 1025 TVSDNKVFVAQNDDGSFTLPLDLADISKFYTTVEDYAGNISYEKV 1069

An alignment of the GAS and GBS proteins is shown below.

Identities = 543/1676 (32%), Positives = 821/1676 (48%), Gaps = 158/1676 (9%)

45 Query: 24 KQRF SIRQYKLGAVSVLLGLTFLFGGITNVAAD--SVINKPSDIAVEQQVKDSPTSI--- 78
KQRF S+RKYK G SVL+G++F + T VAAD S +++P+ QQ T+
Sbjct: 4 KQRFSLRKYKSGTFSVLIGSVFLVM-TTVADELSTMSEPTITNHAQQQAQHLTNTELS 62

50 Query: 79 ANETPTNNTSSALASTAQD-----NLVTKANNSPTETQPVAESH SQATETFPVANQPVE 133
+ E+ + +TS T ++ +LV++ + A + ++ A+ P
Sbjct: 63 SAESKSQDTSQITLKTNREKEQSQDLVSEPTTTELADTDAASMAN TGSDATQKSASLEPPV 122

55 Query: 134 STQEVSKTPLTKQ--NLAVKSTPAISKETPQNID-SNKIITVPKVMNTGYKGEGTVVAI- 189
+T +V TK + K + ID +++ + + V K + ++A
Sbjct: 123 NT-DVHDWVKTKGAWDKGYKGQKVVAVIDTGIDPAHQSMRISDVSTAKVKS EDMLARQ 181

60 Query: 190 -----IDSGLDIN-----HDALQLNDSTKAK-----YQNEQQMNAKAKAGINYGKW 231
I+ G IN H+ ++ +D+ K ++N + A+ KA I K
Sbjct: 182 KAAGINYG SWINDKVVF AHNYVENS DNKENQFEDFEDWENFEFDAAEAPKA- IKKHKI 240

65 Query: 232 YN-----NKVIFGHNYVDVNTLKEVKSTSHGMHVT SIATANPSKKD-TNEL 277
Y + G + +D + K SHGMHVT I N + T E
Sbjct: 241 YRPQSTQAPKETVIKTEETDGS HDIDWTQTD DDTKYESHGMHVTGIVAGNSKEAAATGER 300

Query: 278 IYGVAPAEQVMFMRVFSDEKRGTPALYVKAIEDAVKLGADSINLSLGGANGSLVNADDR 337
G+APEAQVMFMRVF+++ G+ +L++KAIEDAV LGAD INLSL G ANG+ ++
Sbjct: 301 FLGIAPEAQVMFMRV FANDIMGSAESLFIKAIEDAVALGADVINLSLGTANGAQLSGSKP 360

-2302-

5
Query: 338 L I K A L E M A R L A G V S V V I A A G N D G T F G S G A S K P S A L Y P D Y G L V G S P S T A R E A I S V A S Y N N T 397
L++A+E A+ AGVSVV+AAGN+ +GS P A PDYGLVGSPST R SVA+ N+
Sbjct: 361 L M E A I E K A K K A G V S V V A A G N E R V Y G S D H D D P L A T N P D Y G L V G S P S T G R T P T S V A A I N S K 420

10
Query: 398 T L V N K V F N I I G L E N N R N L N N G L A A Y A --- D P K V S D K T F E V G K Q Y D Y V F V G K G N D N D Y K D K 454
++ ++ + L E N + L N + G A Y + D K + K + + + V + D Y +
Sbjct: 421 W V I Q R L M T V K E L E N R A D L N H G K A I Y S E S V D F K D I K D S L G Y D K S H Q F A Y V K E S T D A G Y N A Q 480

15
Query: 455 T L N G K I A L I E R G - D I T F T K K V V N A I N H G A V G A I I F N N K A G E A N L T M S L D P E A S A I P A I F T 513
+ G K I A L I E R + T + + + A H G A + G + I F N N K G + + N + M L I P + F
Sbjct: 481 D V K G K I A L I E R D P N K T Y D E M I A L A K K H G A L G V L I F N N K P G Q S N R S M R L T A N G M G I P S A F I 540

20
Query: 514 Q K E F G D V L A K N N Y K --- I V F N N I K N K Q A N P N A G V L S D F S S W G L T A D G Q L K P D L S A P G G S 569
E F G + + + N + F + + + K + + + F S + W G L T + D G L K P D + + A P G G
Sbjct: 541 S H E F G K A M S Q L N G N G T G S L E F D S V V S K A P S Q K G N E M N H F S N W G L T S D G Y L K P D I T A P G G D 600

25
Query: 570 I Y A A I N D N E Y D M M S G T S M A S P H A G A T A L V K Q Y L L K E H P E L K K G D I E R T V K Y L L M S T A K A 629
I Y + N D N Y + G T S M A S P + A G A + L V K Q Y L K P L K I V K L L M S A +
Sbjct: 601 I Y S T Y N D N H Y G S Q T G T S M A S P Q I A G A S L L V K Q Y L E K T Q P N L P K E K I A D I V K N L L M S N A Q I 660

30
Query: 630 H L N K D T G A Y T S P R Q Q G A G I I D V A A A V Q T G L Y L T G G E N N Y G S V T L G N I K D K I S F D V T V H N I 689
H + N + T T S P R Q Q G A G + + + + A V + G L Y + T G + + N Y G S + + L G N I D + + F D V T V H N +
Sbjct: 661 H V N P E T K T T T S P R Q Q G A G L L N I D G A V T S G L Y V T G - K D N Y G S I S L G N I T D T M T F D V T V H N L 719

35
Query: 690 N K V A K D L H Y T T Y L N T D Q V - - K D G F V T L A P Q Q L G T F T G K T I R I E P G Q T Q T I T I D I D V S K Y H 747
+ K L Y T L T D V + G T L L T + G + + T + + + D V S + +
Sbjct: 720 S N K D K T L R Y D T E L L T D H V D P Q K G R F T L T S H S L K T Y Q G G E V T V P A N G K V T V R V T M D V S Q F T 779

40
Query: 748 D M L K K V M P N G Y F L E G Y V R F T D P V D G G - E V L S I P Y V G F K G E F Q N L E V L E K S I Y K L V A N K E K 806
L K M P N G Y + L E G + V R F D D + + I P + V G F K G + F + N L V E + S I Y + L + +
Sbjct: 780 K E L T K Q M P N G Y Y L E G F V R F R D S Q D D Q L N R V N I P F V G F K G Q F E N L A V A E S I Y R L K S Q G K T 839

45
Query: 807 G F Y F Q P K - Q T N E V P G S E D Y T A L M T T S S E P I Y S T D G T S P I Q L K A L G S Y K S I D G K W I L Q L D Q 865
G F Y F + + + + + T L + T S E S T S L L G + + K + D G K + I L + +
Sbjct: 840 G F Y F D E S G P K D D I Y V G K H F T G L V T L G S E T N V S T K T I S D N G L H T L G T F K N A D G K F I L E K N A 899

50
Query: 866 K G Q P H L A I S P N D D Q N D A V A V K G V F L R N F N N L R A K V Y R A D D V N L Q K P L W S A P Q - A G D K N 924
+ G P L A I S P N D N Q D A K G V F L R + L + A V Y A D + P L W V S G D K N
Sbjct: 900 Q G N P V L A I S P N G D N N Q D F A A F K G V F L R K Y Q G L K A S V Y H A S D K E H K N P L W V S P E S F K G D K N 959

55
Query: 925 Y Y S G N T E N P K S T F L Y D T E W K G T T D G I P L E D G K Y K Y V L T Y Y S D V P G S K P Q Q M V F D I T L D R 984
+ S + K S T L T + G + G L D G Y Y V + + Y Y D V G + K Q + M F D + L D R
Sbjct: 960 F N S - D I R F A K S T T L L G T A F S G K S L T G A E L P D G H Y H Y V S Y Y P D V V G A K R Q E M T F D M I L D R 1018

60
Query: 985 Q A P T L T T A T Y D K D R R I F K A R P A V E H G E S G I F R E Q V F Y L K D K D G H Y N S V L R Q Q G E D G I L V 1044
Q P L + A T + D + F K P + G + G + + V F Y L + + K D + V + V
Sbjct: 1019 Q K P V L S Q A T F D P E T N R F K P E P L K D R G L A G V R K D S V F Y L E R - K D N K P Y T V T I N D S Y K V S V 1077

65
Query: 1045 E D N K V F I K Q E K D G S F I L P K E V N D F S H V Y T V E D Y A G N L V S A K L E D L I N I G N K N G L V N V K V 1104
E D N K F + + + + D G S F I L P + Y Y V E D + A G N + A K L D + + + K +
Sbjct: 1078 E D N K T F V E R Q A D G S F I L P L D K A K L G D F Y M V E D F A G N V A I A K L G D H L P Q T L G K T P I K L K L 1137

Query: 1105 F S P E L N S N V D I D F S Y S V K D D K G N I I K K Q - - - - - H H G K D L N L L K L P F G T Y T F D L F L Y D E E 1158
+ + + + + Q H + + L D F + E
Sbjct: 1138 T D G N Y Q T K E T L K D N L E M T Q S D T G L V T N Q A Q L A V V H R N Q P S Q L T - - - - K M N Q D F F I S P N E 1193

Query: 1159 R A N L I S P K S V T V T I S E K D S L K D V L F K V N L L K K A A L L V E F D K L L P - - - - - K G A T V Q L V T K T 1213
N K K + + + + L V N + K + K P G A + V + T
Sbjct: 1194 D G N - - - - K D F V A F K G L K N N V Y N D L - T V N V Y A K D - - - - D H Q K Q T P I W S S Q A G A S V S A I E S T 1244

Query: 1214 N T V D L P K A T Y S P T D Y G K N I P V G D Y R L N V T L P S G Y S T L E N L D D L L V S V K E D Q V N L T - - - K L 1271
A Y T G + G D Y + V T + E + + S V + + + T +
Sbjct: 1245 - - - - - A W Y G I T A R G S K V M P G D Y Q Y V V T Y R D E H G K - E H Q K Q Y T I S V N D K K P M I T Q G R F 1295

Query: 1272 T L I N K - - - - A P L I N A L A E Q T D I I T Q P V F Y N A G T H L K N N Y L A N L E K A Q T L I K N R V E Q T S I D 1327
I N P + + I + + V F Y A K N + + + + I T D
Sbjct: 1296 D T I N G V D H F T P D K T K A L D S S G I V R E E V F Y L A - - - K K N G R K F D V T E G K D G I - - - - - T V S D 1346

-2303-

Query: 1328 NATAALRESRQALNGKETDTSLLAKAILAETEIKGNVQFVNASPL----SQSTYIN---- 1379
 N + + + + D L+ + GN F L +N
 Sbjct: 1347 NKVYIPKNPDGSGYTISKRDGVTLSDYLLVEDRAGNVSFATLRDLKAVGKDKAVVNFGLD 1406

Query: 1380 -QVQLAKNLLQKPNVTQSEVDKALENLDAKNQNLNGHETDYS--GLHHMIIKANVLKQTS 1436
 V K ++ + + K +ENL+ N N Y + + N K S
 Sbjct: 1407 LPVPEDKQIVNFTYLVRDADGKPIENLEYNNNSGNSLILPYGKYTVELLTYDTNAAKLES 1466

Query: 1437 SKYQNASQFAKENYNNLIKKAEALLSNR-----QATQAQVEELINQIKATEQEL- 1485
 K + + A N+ + K +L +++ + ++ ++ +Q+ EQ L
 Sbjct: 1467 DKIVSFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGSRVSLKTAQDQLIPLEQSLY 1526

Query: 1486 -----DGRDRVSSAENYSQSLNDNDLSLNTTTPINPPNPQALIFKKGMTKES----- 1531
 +G V + + N +NT P N ++ + K G +S
 Sbjct: 1527 VPKAYGKTIVQEGTYEVVVSPLPKGYRIEGNTKVNTLP-NEVHELRLVVKVGSDSTGDH 1585

Query: 1532 -----EVAQKRVLGVTSTQDNQKVKTNKLPKTGESTPKITYTILLFSLMLGLATI 1582
 +Q T LP TGE K+ + + L +LGL +
 Sbjct: 1586 KVMSKNNSQALTASATPTKSTTSATAKALPSTGE---KMGLKLRLVGLVLLGLTCV 1638

SEQ ID 8964 (GBS92) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 2; MW 48kDa).

GBS92-His was purified as shown in Figure 199, lane 9 .

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2038

A DNA sequence (GBSx2149) was identified in *Sagalactiae* <SEQ ID 6299> which encodes the amino acid sequence <SEQ ID 6300>. This protein is predicted to be AzlC family protein. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.80	Transmembrane	212 - 228 (196 - 230)
INTEGRAL	Likelihood = -7.27	Transmembrane	167 - 183 (159 - 185)
INTEGRAL	Likelihood = -5.68	Transmembrane	189 - 205 (188 - 210)
INTEGRAL	Likelihood = -2.28	Transmembrane	17 - 33 (13 - 34)
INTEGRAL	Likelihood = -1.06	Transmembrane	135 - 151 (135 - 151)
INTEGRAL	Likelihood = -1.01	Transmembrane	61 - 77 (60 - 77)

----- Final Results -----

bacterial membrane	---	Certainty=0.4121(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 10235> which encodes amino acid sequence <SEQ ID 10236> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF10212 GB:AE001921 AzlC family protein [Deinococcus radiodurans]
 Identities = 72/224 (32%), Positives = 117/224 (52%), Gaps = 8/224 (3%)

Query: 6 FKEGVKDALPTALGYISIGLAFGIVASASDLSAIEVGLMSALVYGGSAQFAMCALLLAKA 65
 F +G + +P LG + LA+ + A A+ LS + LMS + G++QFA L A A
 Sbjct: 7 FWQGFRLVPLWLGTVPFALAYAVTARAAGLSVGDTCLMSLTTFAGASQFAAAGLFGAHA 66

Query: 66 DLMTITMTVFLVNLRLNMLSLHATTIFKSAHLMNQLAIGTLITDES YGV-LLGEALHHKV 124
 ++I +T FL+N R++L L + L ++ +TDE+YGV ++ A

-2304-

Sbjct: 67 GGLSIVLTFTFLNARHLLYGLSLARELRIT-LPQRVVAQFLTDEAYGVAVVSGARLPGG 125

Query: 125 VSPSWMHGNNVMSYLTWVISTIIGTLLGSTIPNPEMFGLDFALVAMFIGLFVFLFGMLS 184
 ++ +++ G + YL+W +ST++G L GS +P PE G+ F+GL V ++

Sbjct: 126 LTFAPLLGAELSLYLSWNVSTLLGALAGSVLPPEQLGVGVVFLAFLGLLV----PLVV 181

Query: 185 DGKRLVVYVLASVGLSYFLLATFLSGALSVLLATVVGCSVGVVL 228
 D RL + V + GL + L+ L G L +LLA V G +G L

Sbjct: 182 D--RLSLLVALAAGLGGWALSRVLPGGVLVLLAGVGGALLGAAL 223

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2039

- 15 A DNA sequence (GBSx2150) was identified in *S.agalactiae* <SEQ ID 6301> which encodes the amino acid sequence <SEQ ID 6302>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3794(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2040

- 30 A DNA sequence (GBSx2151) was identified in *S.agalactiae* <SEQ ID 6303> which encodes the amino acid sequence <SEQ ID 6304>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence

- 35 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.5087(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 40 A related GBS nucleic acid sequence <SEQ ID 10233> which encodes amino acid sequence <SEQ ID 10234> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04157 GB:AP001508 homosystein methyl transferase [Bacillus halodurans]
 Identities = 397/751 (52%), Positives = 519/751 (68%), Gaps = 14/751 (1%)

Query: 10 SNLGYPRLGGEQREWKAIEAFWAGNLEQKDLEKQLKQLRINHLKKQKEAGIDLIPVGDFS 69

SNLGYPR+GE REWK+A+E+FWA + ++ L +K+LR+NHL+ Q+E +DLIPVGDF+

Sbjct: 4 SNLGYPRIGENREWKKALESFWANDITTEEQLLATMKELRLNHLRVQQEQEVLDLIPVGDF 63

Query: 70 CYDHVLDLSFQFNVIPKRFDEY--ERNLDLYFAIARGDKDNVASSMKKWFNTNYHYIVPE 127
 YDHVLD++ F +IPKRF + L YFA+ARG K+ A M KW+NTNYHYIVPE

-2305-

Sbjct: 64 LYDHLVDMAMVFGIIPKRFLQQGDTPTLSTYFAMARGSKNAQACEMTKWYNTNYHYIVPE 123

Query: 128 WEVETKPHLQNNYLLDLYLEAREVVGDKAKPVITGPITYVSLSSGIVD--FEATVQRLLP 185
+ P L N L+ YLEA+ +G KPVI GP ++V L+ G + + T+Q LLP

5 Sbjct: 124 LH-DAAPRLTKNAPLEAYLEAKNELGIDGKPVILGPYSFVKLAKGYEEDKLQETIQSLLP 182

Query: 186 LYKQVFDLIDAGATYIQIDEPIFVTDEGELLVDIAKSVDFFAREVPQAHFIFQTYFES 245
LY QV Q+L+DAGA IQ+DEP VT + + +Y+ + A QTYF++

10 Sbjct: 183 LYIQVIQELVDAGARSIQVDEPSLVTSSISAREMALVTRIYEQINEAIADAPLFLQTYFDA 242

Query: 246 AVCLDKLSKLPVTGFGGLDFIHGRAENLA AVKQ-GLFREKELFAGIVNGRNIWAVNLEETL 304
+++ LPV G GLDF+HG A+NL A++ G +K L AGI++GRNIW NL E

Sbjct: 243 VTFYEEVVS L P V K G I G L D F V H G G A K N L E A L R T F G F P E D K V L A A G I I D G R N I W I S N L R E R H 302

15 Query: 305 ALL EE I G P F V K -- R L T L Q P S S L L H V P V T T K Y E T H L D P V L K N G L S F A D E K L K E L L L A S A 362
L+ ++ V R L L Q P S S L L H V P V T T K E L D P L L L + F A + E K L E L L

Sbjct: 303 ELVHQLEQHVAKDRLVLQPSCSLHVPVTTKREEKLDPTLLGVLAFA NEKLT ELHTLKQL 362

20 Query: 363 FDGNKTKGYHEALS R --- F S A I Q A A D F R H V A L E S L - A E V K L E R S P Y K L R Q A L Q A E K L Q L 417
GN+ + EAL +AL+ + +R A S E K + R+ LQ EK QL

Sbjct: 363 AAGNEAE-VKEALEANDDALAALEKSGWRSGAATSHNLENKKRPQSFNERRPLQEEKWQL 421

Query: 418 P I L P T T T I G S F P Q S P E I R K K R L A W K R G N L S D S D Y K D F I K T E I R R W I A I Q E D L D L D V L V H G 477
P+LPTTTIGSFQP+ ++R+ R W++G LS +Y+ +K+ I +WI IQE+L LDVLVHG

25 Sbjct: 422 P L L P T T T I G S F P Q T K D V R R T R S L W R K G E L S T V E Y E R T M K S Y I E K W I N I Q E E L G L D V L V H G 481

Query: 478 E F E R V D M V E F F G Q K L A G F T T T K L G W V Q S Y G S R A V K P P I I Y G D V K H I Q P L S L E E T V Y A Q S L 537
EFER DMVEFFG+KL GF T GWVQSYGSR VKPPIIYG+V +P+++ ETVYAQSL

30 Sbjct: 482 E F E R N D M V E F F G E K L D G F A T A N G W V Q S Y G S R C V K P P I I Y G N V S F T E P M T V A E T V Y A Q S L 541

Query: 538 T K K P V K G M L T G P I T I T N W S F E R D D I S R S D L F N Q I A L A I K D E I Q L L E Q S G I A I I Q V D E A A L 597
T KPVKGMLTGP+TI NWSF RDD+ + + +QIA A+ E+ LE++GI +IQ+DE A+

Sbjct: 542 T D K P V K G M L T G P V T I L N W S F V R D D L P L T V I A H Q I A E A L T H E V T A L E E A G I E M I Q I D E P A I 601

35 Query: 598 R E G L P L R Q Q K Q A Y L D D A V A A F K I A T S S V K D E T Q I H T H M C Y S K F D E I I D S I R A L D A D V I S 657
REGLPL+ + QQ YLD AV+AF+ + + VK TQIHTHCYS+F E+I++I LDADVIS

Sbjct: 602 R E G L P L K A E D Q Q E Y L D W A V S A F R A S C A H V K A T T Q I H T H M C Y S E F H E M I E A I D D L D A D V I S 661

40 Query: 658 I E T S R S H G D I I E S F E T A V Y P L G I G L G V Y D I H S P R I P T K E E I I V N I Q R S L K C L S K E Q F W V N 717
IETSRSHG++I +FE Y GIGLGVDIHSR+P++EE++ I+R+L L FWN

Sbjct: 662 I E T S R S H G E M I S A F E K T T Y E K G I G L G V Y D I H S P R V P S E E E M L N V I R R A L T V L P A S L F W V N 721

Query: 718 P D C G L K T R E A E T I A A L E V L V S A T K E V R Q Q L 748
PDCGLKTR E ET+AAL+ +V+A + R++L

45 Sbjct: 722 P D C G L K T R A E K E T V A A L K N M V A A A A R E E L 752

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

50 Example 2041

A DNA sequence (GBSx2152) was identified in *Sagalactiae* <SEQ ID 6305> which encodes the amino acid sequence <SEQ ID 6306>. This protein is predicted to be methH. Analysis of this protein sequence reveals the following:

Possible site: 20
55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm ---	Certainty=0.0753(Affirmative)	< succ>
bacterial membrane ---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside ---	Certainty=0.0000(Not Clear)	< succ>

60

-2306-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05348 GB:AP001512 unknown conserved protein [Bacillus halodurans]
Identities = 301/610 (49%), Positives = 437/610 (71%), Gaps = 9/610 (1%)

```

5   Query: 1   MSKFLEKLKTDILVADGAMGTLTYGLDTCHESYNVTHPEKVLAIHQAYIEAGADVIQT 60
      M+ +E LKT+ILV DGAMGTLTY G+D C E NVT PEK++A H AY+EAGADVIQT
      Sbjct: 1   MTNLVEALKTNILVGDGAMGTLTYEQGIDRCFEELNVTDEPKIVAHHVAYVEAGADVIQT 60

10  Query: 61  NTYGAQRHRLKNYGLDQVVSINQAAVNIAHQATLGKETFILGTVGGRSQRQCDELTDN 120
      NTY A R +L Y L+DQV+ IN+AAV +A +A +ETF+LGT+GG RS + ++ +
      Sbjct: 61  NTYAANRMKLAKYQLDDQVLEINRAAVRLARKAAK-QETFVLGTIGGIRSVQFEEVEIQE 119

      Query: 121 IVEETLEQVEALLATGQLDGLLFETYDYDIEETITVLKIVREMTDLPITNISLHEAGVTS 180
      + + LEQ++AL++ G +DGLL ET+YD+EE + + R +TDLP+I ++S+ E GV
15  Sbjct: 120 VQDVFLQMKALVSEG-VDGLLETFYDLEEAKLAVSLARSLTDLPVIAHLSIAEIGVLQ 178

      Query: 181 NGKPIVEALSQVLMLGADVIGLNCHLGPYHMIQSLKQVPLFAQSYSVYPNASQLSLDGE 240
      GK + EA ++L LGAD++G+NC +GPY M++SL+ V L ++Y S YPNAS D
      Sbjct: 179 GKKLLEEAFAELEGLGADLVGINCRMGFPYQMLRSLETVQLLDRAYYSAYPNASLP--DYS 236

20  Query: 241 NSQYQFSQNSEYFGKSAELLVAEGVRLIGGCCGCTTPDHIRAVKRSIRGLKPIERKVVTP 300
      + + + N EYF + + V +GVRL+GGCCGCTTP+H+RA + ++GLKP+ K V
      Sbjct: 237 DGRLYYHSNPEYFYEMGKRFVQGVRLIGGCCGCTTPPEHVRAFVKVGLKPVVSKPVR-- 294

25  Query: 301 IPVKDFVRRIRRT---DTLVKVKKEVTIIAELDPPKHLDIVQFQKAIRAIDQKGIAAIT 357
      + +K+ + + + L +KVKK+ +II ELDPPK+L I +F + A+ G+ A+T
      Sbjct: 295 LEIKETLSSTGQKTAREPLAEKVKKQPSIIVELDPPKNLAIDRFVEGAAALKNAGVDAVT 354

30  Query: 358 LADNSLSNTRICNLSIASLLKDEISTPFLHIAICRDHNLIGLQSRLLGMELLGFNHILAI 417
      +ADNSL++ R+ NL++ ++++ ++ L+H+ CRD NLIGLQS L+G+ LG +LAI
      Sbjct: 355 MADNSLASPRVDNALGAIQQQVGARPLVHVTCTDRNLIGLQSHLMGLHALGMTDLAI 414

      Query: 418 TGDPTKLGDFPGATSVYDVTSTFKLLSLIKQLNQGLSYSGASLRRPTDFTVAAAFNPNVKN 477
      TGDPTK+GDFPGATSVYDVTSTF+L+SLIKQLN+G+S+SG L + +F+V AAFNPNV++
35  Sbjct: 415 TGDPTKVGDFPGATSVYDVTSTFQLISLIKQLNEGISFSGKELGQKANFSVGAAFNPNVRH 474

      Query: 478 LTRTVKLEIEKKVASGADYFMTQPIFDHSLVKELADLTKTVEQFFFIGIMPITSYNNAVFL 537
      L R V+ +EKK+ +GADYFMTQPI++ ++++ + TK +E+P +IGIMP+ + NA FL
      Sbjct: 475 LERAVQRMKKIEAGADYFMTQPIYNEKQIEDIYEATKHIEKPIYIGIMPLINGRNAEFL 534

40  Query: 538 HNEVPGIKLSSESFLSALEKVKDDKEACLTALNESKSLIDEALNYFNGIYLIPTFLRYDL 597
      HNEVPGIKL++ + + +D++ L +KSL+D A +YFNGIYLIPTFLRY +
      Sbjct: 535 HNEVPGIKLTDQIRERMARAGEDRQKGEREGLAIKSLLDVATHYFNGIYLIPTFLRYGM 594

45  Query: 598 TLELIDYIQK 607
      T++L Y+++
      Sbjct: 595 TVDLTHYVKE 604

```

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2042

A DNA sequence (GBSx2153) was identified in *S.agalactiae* <SEQ ID 6307> which encodes the amino acid sequence <SEQ ID 6308>. Analysis of this protein sequence reveals the following:

```

55  Possible site: 53
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.55    Transmembrane 127 - 143 ( 121 - 147)
      INTEGRAL    Likelihood = -1.44    Transmembrane 157 - 173 ( 155 - 175)

60  ----- Final Results -----
      bacterial membrane --- Certainty=0.4821(Affirmative) < succ>

```

```

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAC01354 GB:AL390975 putative integral membrane protein
[*Streptomyces coelicolor* A3(2)]
Identities = 38/98 (38%), Positives = 59/98 (59%)

Query: 113 RIADDVARFGGSWTFIIIVFSIMAIWMLVNIMKPFGIQFDPPYFILLNLALSTIAAIQAP 172
R+++ VARF G+ FI+ ++ +W++ N+ P G++FD YPFI L L LS A+ AP

Sbjct: 47 RLSERVARFLGTGRFIVWMTVVIIWVWVNVSAPSGLRFDEYYPFIFLTLMLSQASYAAP 106

Query: 173 LIMMSQNRAADYDRLQARNDFNVNKTSELEIRLLHEKI 210
LI+++QNR D DR+ D N+ S + L +I

Sbjct: 107 LILLAQNRODDRRVNLQDRKQNERSTADTEYLTREI 144

55 840 870 900 930 960 990 1020 1050
APLIMMSQNRAADYDRLQARNDFNVNKTSELEIRLLHEKIDHMQDDQFELLEIQKLQTEMLVSLGNQLAQLKQLQK*SF
|||:::|| | ||: | : | : | : |
APLILLAONRQDDRNVNLEQRKONERSIADTEYLTREIAALRIGLGEVATRDWIRSEIODI.VRDLEERONGHHPPDRGV

-2308-

120 130 140 150 160 170 180

SEQ ID 8966 (GBS393) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 3; MW 30.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 177 (lane 4; MW 56kDa) and in Figure 83 (lane 6; MW 56kDa).

GBS393-GST was purified as shown in Figure 217, lane 5.

Example 2043

A DNA sequence (GBSx2154) was identified in *S.galactiae* <SEQ ID 6309> which encodes the amino acid sequence <SEQ ID 6310>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -3.29      Transmembrane 274 - 290 ( 271 - 291)

----- Final Results -----
      bacterial membrane --- Certainty=0.2317(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35508 GB:AE001721 glycerol dehydrogenase [Thermotoga maritima]
Identities = 94/307 (30%), Positives = 157/307 (50%), Gaps = 21/307 (6%)

Query: 63  VYGT DSTQSNIDKLVANPQVQAADAILGFGGKALDTAKMVAKELGKNSFTIPTICSNCS 122
      ++G + +   I++L   + +   D ++G GGGK LDTAK VA +L K   +PTI S +
Sbjct: 62  IFGGECSDBEIERLSGLVE-EETDVVVGIGGGKTLDTAKAVAYKLKKPVVIVPTIASTDA 120

Query: 123 AGTAIAVVYNDHDSFLRYGY-PESPLHIFINTRIIAQAPSKYFWAGIGDGISKAPEVERA 181
      +A++V+Y +   F RY + P +P + ++T I+A+AP+++ AG+GD ++   E E
Sbjct: 121 PCSALSVIITPNGEFKRFLPRNPDVVLVDTEIVAKAPARFLVAGMGDALATWFEAESC 180

Query: 182 TLEAKTNKLPHT-AVLGQVAVALSSKEAFYQFGEQGLKDVEANLASRAVEEI--ALDILIS 238
      +   N       ++   A+A   E   ++G   + VE   + A+E+I A   +L
Sbjct: 181 KQKYAPNMTGRLGSM TAYALARLCYETLLEYGVLAKRVSVEEKSVTPALEKIVEANTLLSG 240

Query: 239 TGYASNVLNQPDFYNSCHAHAFYYGTTAIQRQGEFLHGVVAVFGVLV-LHAYFNELEEL 297
      G+ S               AHA + G T ++   ++LHG VA GVL L       + +
Sbjct: 241 LGFESG-----GLAAAHAIHGLTVLENTHKYLHGEKVAIGVLASLFLTDKPRKMI 291

Query: 298 EKVARFNKSLGLPTTLADVSL---SEKDIPKIVEIAMTNE---YKNTPFDPKMFAQAIL 351
      E+V F + +GLPTTLA++ L   S++D+ K+ E A   NE   + P   K   A+
Sbjct: 292 BEVVSFCEEVGLPTTLAEIGLDGVSDEDLMKVAEKACDKNETIHNEPQPVTSKDVFFALK 351

Query: 352 AADAFGQ 358
      AAD +G+
Sbjct: 352 AADRYGR 358
```

There is also homology to SEQ ID 3078.

SEQ ID 6310 (GBS123) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 7; MW 43.3kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2309-

Example 2044

A DNA sequence (GBSx2155) was identified in *S.agalactiae* <SEQ ID 6311> which encodes the amino acid sequence <SEQ ID 6312>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0974(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6313> which encodes the amino acid sequence <SEQ ID 6314>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2368(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/167 (55%), Positives = 121/167 (72%)

Query: 1 MKIAIIGYSGSGKSTLARKLGNYYNCNVLHLDLSIHFAFNWEERKYDDMIDDVSNMLEKRT 60

+KIAIIG+SGSGKSTLAR LG +Y+C V HLD +HF+ NW+ER DMI D+S L K+

Sbjct: 1 LKIAIIGHSGSGKSTLARFLGQHYHCEVFHLDQLHFSSNWQERSDHDMIADLSTCLLKQD 60

Query: 61 WIIEGNYKKLLYQERLADAEIIFDFNRFNCLWRAFKRYCKFRGKTRPDMANGCPEKLD 120

IIEGNY LY+ER+++AD II+ +F+RF+C++RAFKRY +RGKTRPDMA+ C EK D

Sbjct: 61 LIIEGNYANCLYEERMSEADYIIYVNFSRFHCVYRAFKRYLNYRGKTRPDMADNCQEKFD 120

Query: 121 FEFISWILKDGRSDKQKSNYKQVVEDYPQKIKILKHQRDLDDQYLKEL 167

F+ WIL DGRS Q Y+ VV+ Y K +L +Q+ L Y+ +

Sbjct: 121 VAFVKWILLDGRSRNQLKKYQSVVQKYSHKTIIVLTNQQLSHYMNTI 167

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2045

A DNA sequence (GBSx2156) was identified in *S.agalactiae* <SEQ ID 6315> which encodes the amino acid sequence <SEQ ID 6316>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3874(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA41941 GB:X59250 initiation factor IF-1 [Lactococcus lactis]

Identities = 62/72 (86%), Positives = 70/72 (97%)

Query: 1 MAKEDVIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKQNYIRILVGDRVTVEMSPY 60

MAK+DVIE++GKVV+TMPNAMFTVELENGHQ+LAT+SGKIRKQNYIRIL GD+V VE+SPY

Sbjct: 1 MAKDDVIEVDGKVVDTMPNAMFTVELENGHQVLATISGKIRKQNYIRILPGDKVQVELSPY 60

-2310-

Query: 61 DLTRGRITYRFK 72
DLTRGRITYRFK
Sbjct: 61 DLTRGRITYRFK 72

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6317> which encodes the amino acid sequence <SEQ ID 6318>. Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3253(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 67/67 (100%), Positives = 67/67 (100%)

Query: 6 VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVMSPYDLTRG 65
VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVMSPYDLTRG
Sbjct: 1 VIEIEGKVVETMPNAMFTVELENGHQILATVSGKIRKKNYIRILVGDRVTVMSPYDLTRG 60

Query: 66 RITYRFK 72
RITYRFK

Sbjct: 61 RITYRFK 67

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2046

A DNA sequence (GBSx2157) was identified in *S.agalactiae* <SEQ ID 6319> which encodes the amino acid sequence <SEQ ID 6320>. This protein is predicted to be adenylate kinase (adk). Analysis of this protein sequence reveals the following:

Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA41940 GB:X59250 adenylate kinase [Lactococcus lactis]
Identities = 146/214 (68%), Positives = 170/214 (79%), Gaps = 6/214 (2%)

Query: 1 MNLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMRLAKSYIDKGELVP 60
MNLLIMGLPGAGKGTQA IV+ +GV HISTGDMFRAAM N+TEMG+LAKS+IDKGELVP
Sbjct: 1 MNLLIMGLPGAGKGTQAEFIVKNYGVNHISTGDMFRAAMKNETEMGKLAKSFIDKGELVP 60

Query: 61 DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC 120
DEVTNGIVKERLA+DDI GFLLDGYPRTI+QAHALD LEELG++LD V+NI V+P+
Sbjct: 61 DEVTNGIVKERLAQDDIKASGFLLDGYPRTIDQAHALDTMLEELGIKLDVAVNVIVNPN 120

Query: 121 LIERLSGRIINRKTGETFHKVFNPV-----DYKEEDYYQREDDKPETVKRRLDVNIAQ 174
L++RLSGR I R G T+HK+FNP D YQR DD PETVK RLDVNI +
Sbjct: 121 LVDRLSGRYICRNCGATYHKIFNPPTKVEGTCDCVCGSHDLYQRADVPETVKRRLDVNIKE 180

Query: 175 GEPILEHYRKLGLVTDIEGNQEITEVFADVEKAL 208
PI+EHY +LGLV +IEG QEI++V D++K L

-2311-

Sbjct: 181 SAPIIEHYTELGLVKNIEGEQEISQVTDIDKKVL 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6321> which encodes the amino acid sequence <SEQ ID 6322>. Analysis of this protein sequence reveals the following:

5 Possible site: 17
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 208/212 (98%), Positives = 212/212 (99%)

15 Query: 1 MNLLIMGLPGAGKGTQAAKIVEEFGVAHISTGDMFRAAMANQTEMGRILAKSYIDKGELVP 60
 MNLLIMGLPGAGKGTQAAKIVEEFG+AHISTGDMFRAAMANQTEMGRILAKSYIDKGELVP
 Sbjct: 1 MNLLIMGLPGAGKGTQAAKIVEEFGIAHISTGDMFRAAMANQTEMGRILAKSYIDKGELVP 60

20 Query: 61 DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC 120
 DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC
 Sbjct: 61 DEVTNGIVKERLAEDDIAEKGFLLDGYPRTIEQAHALDATLEELGLRLDGVINIKVDPSC 120

25 Query: 121 LIERLSGRIINRKTGETFHKVFNPVDYKEEDYYQREDDKPETVKRRLDVNIAQGEPIL 180
 L+ERLSGRIINRKTGETFHKVFNPVDYKEEDYYQREDDKPETVKRRLDVN+AQGEPIL
 Sbjct: 121 LVERLSGRIINRKTGETFHKVFNPVDYKEEDYYQREDDKPETVKRRLDVNMAQGEPIL 180

 Query: 181 HYRKLGLVTDIEGNQEITEVFADVEKALLELK 212
 HYRKLGLVTDIEGNQEIT+VFADVEKALLELK
 30 Sbjct: 181 HYRKLGLVTDIEGNQEITDVFADVEKALLELK 212

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

35 A related GBS gene <SEQ ID 8967> and protein <SEQ ID 8968> were also identified. Analysis of this protein sequence reveals the following:

 Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: -1.04
 GvH: Signal Score (-7.5): -1.08
 Possible site: 17

40 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 0 value: 6.79 threshold: 0.0
 PERIPHERAL Likelihood = 6.79 106
 modified ALOM score: -1.86

45 *** Reasoning Step: 3

 ----- Final Results -----

50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

over 213aa

55 EGAD|8612| adenylate kinase Insert characterized
 SP|P27143|KAD_LACLA ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE). Edit
 characterized
 GP|44074|emb|CAA41940.1||X59250 adenylate kinase Insert characterized
 PIR|S17987|S17987 adenylate kinase (EC 2.7.4.3) - subsp. lactis Insert characterized
 60 PIR|B44812|B44812 adenylate kinase (EC 2.7.4.3) - Insert characterized

Lactococcus lactis

----- Final Results -----

-2313-

bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9467> which encodes amino acid sequence <SEQ ID 9468> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA41939 GB:X59250 SecY protein [Lactococcus lactis]
 Identities = 292/433 (67%), Positives = 361/433 (82%), Gaps = 2/433 (0%)

10 Query: 1 MFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLEQMGELPFLNMLNLV 60
 MF K L++A KVK VR +ILFTIFIL VFR+G HIT PG+NV++L+Q+ +LPFL+M+NLV
 Sbjct: 1 MFFKTLKEAFKVKDVRARILFTIFILFVFRLGAHITAPGVNVQNLQQVADLPFLSMMNLV 60

15 Query: 61 SGNAMRNFSVFSMGVSPYITASIVVQLQMDILPKFVEWGKQGEVGRKLNQATRYISLF 120
 SGNAM+N+S+F+MGVSPYITASI+VQLQMDILPKFVEW KQGE+GRKLNQATRYI+L
 Sbjct: 61 SGNAMQNYSLFAMGVSPYITASIVVQLQMDILPKFVEWSKQGEIGRKLNQATRYITLV 120

20 Query: 121 LAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIGAILTTGSMVVTWLGEQITDKGFGNGV 180
 LA QSIGITAGF +SS+ +V+ PN Q+YL+IG +LTTGSMVVTW+GEQI +KGFG+GV
 Sbjct: 121 LAMAQSIGITAGFQAMSSLNIVQNPWQSYLMIGVLLTTGSMVVTW+GEQINEKGFGSGV 180

25 Query: 181 SMIIFAGIISIPSIAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAIVFFTTFIQQAE 240
 S+IIFAGI+S IPSAI ++Y++ F+NVR S I S+IFV LI++ + I++ TTF+QQAE
 Sbjct: 181 SVIIFAGIVSGIPSAIKSVYDEKFLNVRPSEIPMSWIFVIGLILSAIVIIYVTTFFVQQAE 240

30 Query: 241 YKIPIQYTKLVQGAPTSSYLPLKVNPAIVIPVIFASSITTIPSTIIPFFQ--NGKEIPWL 298
 K+PIQYTKL QGAPTSSYLPL+VNPAGVIPVIFA SITT P+TI+ F Q G + WL
 Sbjct: 241 RKVPIQYTKLTQGAPTSSYLPLRVNPAIVIPVIFAGSITTAPATILQFLQRSQGSNVGWL 300

35 Query: 299 TKLQELLNYQTPVGMIIYAILIILFSFFYTFVQVNPEKTAENLQKNSSYIPSIRPGRETE 358
 + LQ L-Y T GM+ YA+LI+LF+FFY+FVQVNPEK AENLQK SYIPS+RPG+ TE
 Sbjct: 301 STLQNALSYTTWTGMLFYALLLIVLFTFFYSFVQVNPEKMAENLQKQGSYIPSVRPGKGTE 360

40 Query: 419 GYLLKRRYVGFMN 431
 GYLLKR+Y GFM+
 Sbjct: 421 GYLLKRKYAGFMD 433

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3987> which encodes the amino acid sequence <SEQ ID 3988>. Analysis of this protein sequence reveals the following:

45 Possible site: 55
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -14.70	Transmembrane	233 - 249 (226 - 255)
INTEGRAL	Likelihood = -8.12	Transmembrane	330 - 346 (323 - 350)
INTEGRAL	Likelihood = -6.10	Transmembrane	384 - 400 (378 - 403)
50 INTEGRAL	Likelihood = -5.20	Transmembrane	35 - 51 (33 - 56)
INTEGRAL	Likelihood = -4.09	Transmembrane	199 - 215 (195 - 215)
INTEGRAL	Likelihood = -3.56	Transmembrane	167 - 183 (165 - 184)
INTEGRAL	Likelihood = -1.65	Transmembrane	411 - 427 (411 - 428)
INTEGRAL	Likelihood = -1.49	Transmembrane	133 - 149 (133 - 149)
55 INTEGRAL	Likelihood = -0.64	Transmembrane	286 - 302 (285 - 302)

----- Final Results -----

bacterial membrane --- Certainty=0.6880(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

An alignment of the GAS and GBS proteins is shown below.

-2314-

Identities = 377/434 (86%), Positives = 417/434 (95%)

Query: 1 MFLKLLRDALKVKMVRNKILFTIFILLVFRIGTHITVPGINVKSLQMGELPFLNMLNLV 60
 Sbjct: 17 MFLK+L+DALK+K VRNKI FTIFI+LVFRIGTHITVPG+N KSLEQ+ ELPFLNMLNLV 76

Query: 61 SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRKLNQATRYISLF 120
 SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRKLNQATRYISL 136
 Sbjct: 77 SGNAMRNFSVFSMGVSPYITASIVVQLLQMDILPKFVEWGKQGEVGRKLNQATRYISLV 136

Query: 121 LAFVQSIGITAGFNTLSSVALVKTPNVQTYLLIGAILTTGSMVVTWLGEQITDKGFGNGV 180
 LAF QSIGITAGFNTLS+VALVKTP+++TYLLIGA+LTIGS++VTWLGEQITDKGFGNGV 196
 Sbjct: 137 LAFVQSIGITAGFNTLSSVALVKTPDIKTYLLIGALLTTGSMVVTWLGEQITDKGFGNGV 196

Query: 181 SMIIFAGIISSIPSAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAIVFFTTFIQQA 240
 SMIIFAGIISSIPSAI TI ED+FVN++S + +SY+ VGILI+AVLAIVFFTT++QQA 256
 Sbjct: 197 SMIIFAGIISSIPSAIATIREDYFVNVRSSDLHSSYLIVGILIIAVLAIVFFTTVYQQA 256

Query: 241 YKIPIQYTKLVQGAFTSSYLPLKVNPAAGVIPVIFASSITTIPSTIIPFFQNGKEIPWLTK 300
 YKIPIQYTKL+QGAFTSSYLPLKVNPAAGVIPVIFASSITTIPSTIIPF QNG+++PWL + 316
 Sbjct: 257 YKIPIQYTKLVQGAFTSSYLPLKVNPAAGVIPVIFASSITTIPSTIIPFVQNGRDLPLWLN 316

Query: 301 LQELLNYQTPVGMIIYAILIILFSFFYTFVQVNPEKTAENLQKNSSYIPSIRPGRETEEY 360
 LQE+ NYQTPVGM+YA+LIILFSFFYTFVQVNPEKTAENLQKNSSYIPS+RPGRETE++ 376
 Sbjct: 317 LQEIFNYQTPVGMIVYALLIILFSFFYTFVQVNPEKTAENLQKNSSYIPSVRPGRETEQF 376

Query: 361 MSSLLKKLATIGSVFLAFISLLPIAQQALHLSSSIALGGTSLLLILATGIEGMKQLEGY 420
 MS+LLKKLAT+G++FLAFISL PI AQQAL+LSSSIALGGTSLLLILI+TGIEGMKQLEGY 436
 Sbjct: 377 MSALLKKLATVGAIFLAFISLPIAQQALNLSSSIALGGTSLLLILISTGIEGMKQLEGY 436

Query: 421 LLKRRYVGFMMNTE 434
 LLKR+YVGFMMNTE 450
 Sbjct: 437 LLKRRYVGFMMNTAE 450

A related GBS gene <SEQ ID 8969> and protein <SEQ ID 8970> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 6.16
 GvH: Signal Score (-7.5): -4.32
 Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 9 value: -14.01 threshold: 0.0

INTEGRAL	Likelihood = -14.01	Transmembrane	217 - 233 (209 - 240)
INTEGRAL	Likelihood = -9.98	Transmembrane	311 - 327 (307 - 334)
INTEGRAL	Likelihood = -6.16	Transmembrane	369 - 385 (363 - 392)
INTEGRAL	Likelihood = -5.36	Transmembrane	19 - 35 (17 - 40)
INTEGRAL	Likelihood = -3.93	Transmembrane	180 - 196 (179 - 199)
INTEGRAL	Likelihood = -3.03	Transmembrane	395 - 411 (392 - 412)
INTEGRAL	Likelihood = -2.55	Transmembrane	151 - 167 (151 - 168)
INTEGRAL	Likelihood = -2.02	Transmembrane	117 - 133 (117 - 133)
INTEGRAL	Likelihood = -0.64	Transmembrane	270 - 286 (269 - 286)
PERIPHERAL	Likelihood = 0.95		69

modified ALOM score: 3.30

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01657(301 - 1596 of 1902)

EGAD|6545|6344(1 - 434 of 439) preprotein translocase secy subunit {Lactococcus lactis}

SP|P27148|SECY_LACLA PREPROTEIN TRANSLOCASE SECY SUBUNIT. GP|44073|emb|CAA41939.1||X59250

-2315-

```
SecY protein {Lactococcus lactis} PIR|S17985|S17985 preprotein translocase secY -
Lactococcus lactis subsp. lactis
%Match = 46.6
%Identity = 67.0 %Similarity = 84.1
Matches = 290 Mismatches = 68 Conservative Sub.s = 74
```

72 102 132 162 192 222 252 282
HQCKRICSECP*PIKCL*RWY*SNSSCS*RSWNRAC*KIRR*NSW*W*IN*EIVC*SS*IF*IC*SSYHC*RWFNRLSHLI

312 342 372 402 432 462 492 522
NER*LIMFLKLLRDALKVKMVRNKILFTTIFILLVFRIGTHITVPGINVKSLEQMGPPLFNLMLNLVSGNAMRNFVSFMSG
| |: | :: || | : |||||::|:| | || | :|:::|: : |||:|:| |||:|:| |
MFFKTLKEAFKVKDVRARILEFTIFILFVRLGAHITAPGVNVQNLQQVADLPFLSMMNLVSGNAMQNYSLFAMG
10 20 30 40 50 60 70

[illegible]

792 822 852 882 912 942 972 1002
AILTTGSMVVTWLGEQITDKGFNGVSMIIFAGIISSIPSAITTIYEDFFVNVRSSAITNSYIFVGILIVAVLAIVFFT
:|||||::|||:||||:||||:| |||| :::: |:|| | | :|| | ||::: |:|
VLLTTGSMVVWTMGEQINEKGFGSGVSVIIFAGIVSGIPSAIKSVYDEKFLNVRPSEIPMSWIFVGILLSAIVIYVTT

170 180 190 200 210 220 230

1032 1062 1092 1122 1152 1176 1206 1236
 FIQQAEYKIPIQYTKLVQGAPTSSYLPLKVNPAGVIPVIFAGSITTIPSTIIPFQ--NGKEIPWLTKLQELLNYQTVPVG
 |:||| :||||| :||||| :||||| :||||| :||| :||: |:| | : ||: || }:|| |
 FVQQAERKVPIQYTKLTQGAPTSSYLPLRVNPAGVIPVIFAGSITTAPATILQLRSQGSNVGWLSTLQNALSYYTTWTG
 250 260 270 280 290 300 310

[illegible]

1506 1536 1566 1596 1626 1656 1686 1716
 SSSIALGGTSLILILATGIEGМКQLEGYLLKRRVVGFMNTTE*NIG*LCQPSILFFNKSDMLCWIYLLKTK*GDYNESFNY
 : ||||| | : : ||||| | ||| :
 PKIVALGGTSLILIQVAIQAVKQLEGYLLKRYAGFMDNPLETK
 410 420 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2048

A DNA sequence (GBSx2159) was identified in *S. agalactiae* <SEQ ID 6325> which encodes the amino acid sequence <SEQ ID 6326>. This protein is predicted to be 50S ribosomal protein L15 (rplO). Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

```

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5259(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB54021 GB:U96620 ribosomal protein L15 [Staphylococcus aureus]

-2316-

Identities = 116/146 (79%), Positives = 128/146 (87%)

Query: 1 MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60
 Sbjct: 1 MKLHELKPAEGSRKERNRVGRGVATGNGKTSGRGHKGQKARSGGGVRLPGFEGGQLPLFRR 60

Query: 61 MPKRGFSNINAKAYALVNLDQLNVFEDGTETVPVVLKEAGIVRAEKSGVKILNGELTKK 120
 +PKRGF+NIN KEYA+VNLDQLN FEDGTETVP +L E+G+V+ EKSG+KILNG L KK
 Sbjct: 61 LPKRGFTNINRKEYAIVNLDQLNKFEDGTETVPALLVESGVVINEKSGIKILNGSLDKK 120

Query: 121 LSVKAAKFSKSAEAAITAKGGSIEVI 146
 L+VKA KFS SA AI AKGG+ EVI
 Sbjct: 121 LTVKAHKFSASAAEIDAKGGAHEVI 146

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6327> which encodes the amino acid sequence <SEQ ID 6328>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5329(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 135/146 (92%), Positives = 142/146 (96%)

Query: 1 MKLHELKPAEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60
 Sbjct: 1 MKLHELK AEGSRKVRNRVGRGTSSGNGKTSGRGQKGQKARSGGGVRLGFEGGQTPLFRR 60

Query: 61 MPKRGFSNINAKAYALVNLDQLNVFEDGTETVPVVLKEAGIVRAEKSGVKILNGELTKK 120
 +PKRGF+NIN KEYALVNLDQLNVF+DGTEVTP +LK+AGIVRAEKSGVK+LGNGELTKK
 Sbjct: 61 IPKRGFTNINTKEYALVNLDQLNVFDDGTETVPAILKDAGIVRAEKSGVKVLNGELTKK 120

Query: 121 LSVKAAKFSKSAEAAITAKGGSIEVI 146
 L+VKA KFS SA AI AKGG+ EVI
 Sbjct: 121 LTVKAHKFSASAAEIIAKGGSIEVI 146

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2049

A DNA sequence (GBSx2160) was identified in *S.agalactiae* <SEQ ID 6329> which encodes the amino acid sequence <SEQ ID 6330>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1162(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB54020 GB:U96620 ribosomal protein L30 [Staphylococcus aureus]
 Identities = 40/58 (68%), Positives = 46/58 (78%)

Query: 1 MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVNAISHLVTVEE 58
 MA+++ITLT+S IGR QRKTV ALGL K NSSV EDN AIRG +N + HLVTVEE

-2317-

Sbjct: 1 MAKLQITLTSVIGRPETQRKTVEALGLKKTNSVVDNPAIRGQINKVKHLVTVEE 58

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6331> which encodes the amino acid sequence <SEQ ID 6332>. Analysis of this protein sequence reveals the following:

5 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 56/58 (96%), Positives = 57/58 (97%)

 Query: 1 MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVNAISHLVTVEE 58
 MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMV AISHLVTVE+
 Sbjct: 1 MAQIKITLTKSPIGRKPEQRKTVVALGLGKLNSSVVKEDNAAIRGMVTAISHLVTVED 58

20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2050

A DNA sequence (GBSx2161) was identified in *S.agalactiae* <SEQ ID 6333> which encodes the amino acid sequence <SEQ ID 6334>. Analysis of this protein sequence reveals the following:

25 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2051

A DNA sequence (GBSx2162) was identified in *S.agalactiae* <SEQ ID 6335> which encodes the amino acid sequence <SEQ ID 6336>. This protein is predicted to be 30S ribosomal protein S5 (rpsE). Analysis of this protein sequence reveals the following:

40 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.3179(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAA22699 GB:M57621 ribosomal protein S5 [Bacillus
 stearothermophilus]

-2318-

Identities = 119/158 (75%), Positives = 139/158 (87%)

Query: 6 NAVELEERVVAINRVTKVVKGGRRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIRKAVEA 65
 N +ELEERVVA+NRV KVVKGGRRRLRF+ALVVVG+NG VGFGTGKAQEVPEAIRKA+E
 Sbjet: 7 NKLELEERVVAVNRVAKVVKGGRRRLRFSAALVVVGDKNGHVGFGTGKAQEVPEAIRKAIED 66

Query: 66 AKKNMVEVPMVGTTPHEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVADITSK 125
 AKKN++EVP+VGTTPHEV FG +++LKPA EG GV AGG RAV+ELAG++DI SK
 Sbjet: 67 AKKNLIEVPIVGTTPHEVIGHFGAGEIILKPASEGTGVIAGGPARAVLELAGISDILSK 126

Query: 126 SLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDI 163
 S+GSNTPIN+VRAT +GLKQLKRAE+VA LRG +V +L
 Sbjet: 127 SIGSNTPINMVRATFDGLKQLKRAEDVAKLRGKTVEEL 164

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6337> which encodes the amino acid sequence <SEQ ID 6338>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3179(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 158/164 (96%), Positives = 161/164 (97%)

Query: 1 MAFKDNAVELEERVVAINRVTKVVKGGRRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIR 60
 MAFKDNAVELEERVVAINRVTKVVKGGRRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIR
 Sbjet: 1 MAFKDNAVELEERVVAINRVTKVVKGGRRRLRFAALVVVGDRNGRVGFGTGKAQEVPEAIR 60

Query: 61 KAVEAAKKNMVEVPMVGTTPHEVRSEFGGAKVLLKPAVEGAGVAAGGAVRAVIELAGVA 120
 KAVEAAKKNM+EVPMVGTTPHEV + FGGAKVLLKPAVEG+GVAAGGAVRAVIELAGVA
 Sbjet: 61 KAVEAAKKNMIEVPMVGTTPHEVYTNFGGAKVLLKPAVEGSGVAAGGAVRAVIELAGVA 120

Query: 121 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA 164
 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA
 Sbjet: 121 DITSKSLGSNTPINIVRATVEGLKQLKRAEEVAALRGISVSDLA 164

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2052

- A DNA sequence (GBSx2163) was identified in *S.agalactiae* <SEQ ID 6339> which encodes the amino acid sequence <SEQ ID 6340>. This protein is predicted to be 50S ribosomal protein L18 (rplR). Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4488(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 55 A related GBS nucleic acid sequence <SEQ ID 9465> which encodes amino acid sequence <SEQ ID 9466> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-2319-

>GP:AAB06815 GB:L47971 ribosomal protein L18 [Bacillus subtilis]
Identities = 86/120 (71%), Positives = 97/120 (80%), Gaps = 2/120 (1%)

Query: 4 VISKPDKNKIRQKRHRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASASTLD 63
+I+K KN R KRH RVR KLSGTA+RPRLN+FRSN IYQA+IDDV GVTLASASTLD
Sbjct: 1 MITTKSKNAARLKRHARVRAKLSGTAERPRNLNFRSNKHIYAQIIDDVNGVTLASASTLD 60

Query: 64 KE--VSNGTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALADSARENGLKF 121
K+ V + T A VG+LVA+RA KGIS+VVFDRGGYLYHGRVKALAD+ARE GLKF
Sbjct: 61 KDLNVESTGDTSAATKVGEVLAKRAAEKGISDVVFDRGGYLYHGRVKALADAAREAGLKF 120

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6341> which encodes the amino acid sequence <SEQ ID 6342>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4488(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 116/121 (95%), Positives = 120/121 (98%)

Query: 1 MKIVISKPDKNKIRQKRHRVRGKLSGTADRPRLNIFRSNTGIYAQVIDDVAGVTLASAS 60
+KIVISKPDKNKIRQKRHRVRGKLSGTADRPRLN+FRSNTGIYAQVIDDVAGVTLASAS
Sbjct: 1 VKIVISKPDKNKIRQKRHRVRGKLSGTADRPRLNFRSNTGIYAQVIDDVAGVTLASAS 60

Query: 61 TLDKEVSNGTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALADSARENGLKF 121
TLDK+VS GTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALAD+ARENGLKF
Sbjct: 61 TLDKDVSKGTKTEQAVVVGKLVRAERAVAKGISEVVFDRGGYLYHGRVKALADAARENGLKF 121

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2053

A DNA sequence (GBSx2164) was identified in *S.agalactiae* <SEQ ID 6343> which encodes the amino acid sequence <SEQ ID 6344>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1530(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA22700 GB:M57622 ribosomal protein L6 [Bacillus
stearothermophilus]
Identities = 108/178 (60%), Positives = 133/178 (74%)

Query: 1 MSRIGNKVITLPAQVEIINKDNVVTGKPGKQLTREFNKNIGITVEGTEVTITRPNDSKE 60
M R+G K I +PAGV + N VTVKPGKG+LTR F+ ++ ITVEG +TVTRP+D K
Sbjct: 1 MXRVGKKPIEIPAGVTVTVNGNTVTVKPGKGLTRTFHPDMTITVEGNVITVTRPSDEKH 60

Query: 61 MKTIHGTTTRANLNMVVGVSSEGFKKALEMRGVGYRAQLQGSKLVLVSGKSHQDEVEAPEG 120
+ +HGITR+ L NMV GVS+G++KALE+ GVGYRA QG KLVLVSG SH E+E EG
Sbjct: 61 HRALHGTTTSSLANMVEGVSKGEKALELVGVGYRASKQGGKLVLVSGYSHPVEIEPEEG 120

Query: 121 VTFEVPTPTTINVIGINKESVGQTAAYVRSLSPEPYKKGIRYVGEFVRRKEGKTGK 178

-2320-

+ EVP+ T I V G +K+ VG+ AA +R++R PEPYKGKGIRY GE VR KEGKTGK
 Sbjct: 121 LEIEVPSQTKIIVKGADQQRVGELAANIRAVRPPEPYKGKGIRYEGELVRLKEGKTGK 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6345> which encodes the amino acid
 5 sequence <SEQ ID 6346>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1704(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 153/178 (85%), Positives = 166/178 (92%)

Query: 1 MSRIGNKVITLPAGVEIINKDNVTVKGPKGQLTREFNKNIGITVEGTEVTVTRPNDKE 60
 MSRIGNKVIT+PAGVE+ N +NV+TVKGPKG+LTREFNKN I VEGTE+TV RPNDKE
 20 Sbjct: 1 MSRIGNKVITMPAGVELTNNNVITVKGPKGELTREFNKNIEIKVEGTEITVVRPNDKE 60
 Query: 61 MKTTHGTTRANLNNMVGVSEGFKALEMGRGVYRAQLQGSKLVLSVGKSHQDEVEAPEG 120
 MKTTHGTTRANLNNMVGVSEGFKA LEM+GVGYRAQLQG+KLVLSVGKSHQDEVEAPEG
 25 Sbjct: 61 MKTTHGTTRANLNNMVGVSEGFKALEMKGVGYRAQLQGTCLVLSVGKSHQDEVEAPEG 120
 Query: 121 VTFEVPTPTTINVIGINKESVGQTAAYVRSLSPEPYKGKGIRYVGEFVRRKEGKTGK 178
 +TF V PT+I+V GINKE VGQTAAY+RSLRSPEPYKGKGIRYVGE+VR KEGKTGK
 Sbjct: 121 ITFTVANPTSISVEGINKEVVGQTAAYIRSLRSPEPYKGKGIRYVGEYVRLKEGKTGK 178

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 2054

A DNA sequence (GBSx2165) was identified in *S.agalactiae* <SEQ ID 6347> which encodes the amino
 acid sequence <SEQ ID 6348>. This protein is predicted to be 30S ribosomal protein S8 (rpsH). Analysis of
 this protein sequence reveals the following:

35 Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.4356(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAB06813 GB:L47971 ribosomal protein S8 [Bacillus subtilis]
 Identities = 100/132 (75%), Positives = 116/132 (87%)

Query: 1 MVMTPDIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKREGFVKNVEVIEDDKQGII 60
 MVMTPDIAD LTRIRNAN +HE LE+PAS +K+ IA+ILKREGF+++VE +ED KQGII
 50 Sbjct: 1 MVMTPDIADMLTRIRNANMVRHEKLEIPASKLKREIAEILKREGFIRDVEFVEDSKQGII 60
 Query: 61 RVFLKYQNGERVITNLKRISKPLRVYTKHEDMPKVLNGLGIAIVSTSEGLLTDKEARQ 120
 RVFLKYQN ERVIT LKRISKPLRVY K ++P+VLNGLGIAI+STS+G+LTDKEAR
 Sbjct: 61 RVFLKYQNNERVITGLKRISKPLRVYAKSNEVPRVLNGLGIAIISTSQGVLTDEARA 120
 55 Query: 121 KNIGGEVLAYIW 132
 K GGEVLAY+W
 Sbjct: 121 KQAGGEVLAYVW 132

-2321-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6349> which encodes the amino acid sequence <SEQ ID 6350>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4327(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 122/132 (92%), Positives = 129/132 (97%)

Query: 1 MVMTDPIADFLTRIRNANQAKHEVLEVPASNIKKGIADILKREGFVKNEVEIEDDKQGII 60
 Sbjct: 1 MVMTDPIADFLTRIRNANQ KHEVLEVPASNIKKGIA+ILKREGFVKNEVEIEDDKQGII 60

Query: 61 RVFLKYGQNGERVITNLKRISKPLRVYTKHEDMPKVLNGLGIAIVSTSEGLLTDKEARQ 120
 Sbjct: 61 RVFLKYG+NGERVITNLKRISKPLRVY K +DMPKVLNGLGIAI+STSEGLLTDKEARQ 120

Query: 121 KNIGGEVLAYIW 132
 KN+GGEV+AY+W
 Sbjct: 121 KNVGGEVIAYVW 132

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2055

A DNA sequence (GBSx2166) was identified in *S.agalactiae* <SEQ ID 6351> which encodes the amino acid sequence <SEQ ID 6352>. This protein is predicted to be ribosomal protein S14 (rpsN). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11905 GB:Z99104 ribosomal protein S14 [Bacillus subtilis]
 Identities = 47/61 (77%), Positives = 53/61 (86%)

Query: 1 MAKKSMIAKNKRPAKFSTQAYTRCEKCGRPHSVYRKQFQLCRVCFRDLAYKGQVPGVTKAS 60
 MAKKSMIAK +R KF Q YTRCE+CGRPHSV RKF+LCR+CFR+LAYKGQ+PGV KAS
 Sbjct: 1 MAKKSMIAKQORTPKFKVQEYTRCERCGRPHSVIRKFKLCRICFRELAYKGQIPGVKKAS 60

Query: 61 W 61
 W
 Sbjct: 61 W 61

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6353> which encodes the amino acid sequence <SEQ ID 6354>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2322-

bacterial cytoplasm --- Certainty=0.4747(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 55/61 (90%), Positives = 59/61 (96%)

Query: 1 MAKKSMIKKNRPKAKFSTQAYTRCEKCGRPHSVYRKFKLCRVCFRDLAYKGQVPGVTKAS 60

+AKKSMIKKNRPKAK STQAYTRCEKCGRPHSVYRKFKLCRVCFR+LAYKGQ+PGV KAS

10 Sbjct: 1 LAKKSMIKKNRPKAKHSTQAYTRCEKCGRPHSVYRKFKLCRVCFRELAYKGQIPGVVKAS 60

Query: 61 W 61

W

Sbjct: 61 W 61

15

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2056

20 A DNA sequence (GBSx2167) was identified in *S.agalactiae* <SEQ ID 6355> which encodes the amino acid sequence <SEQ ID 6356>. This protein is predicted to be 50S ribosomal protein L5 (rpIE). Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

25

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1845(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03865 GB:AP001507 ribosomal protein L5 (BL6) [Bacillus halodurans]

Identities = 143/178 (80%), Positives = 162/178 (90%)

Query: 3 NRLKEKYTNEVVPALTEKFNYSSVMAVPKVEKIVLNMGVGDAVSNAKMLEKAAAEALALIS 62

35

NRLKEKY E+VP+LTEKFNYSSVMAVPK+EKIV+NMGVGDAV NAK L+KA EL I+

Sbjct: 2 NRLKEKYQKEIVPSLTEKFNYSSVMAVPKLEKIVVNMGVGDAVQNAKALDKAVEELTEIT 61

Query: 63 GQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPTKS 122

40

GQKP+ITKAKKSIAGF+LREG+ IGAKVTLRGERMYEFLDKL+SVSLPRVRDF G+ K+

Sbjct: 62 GQKPIITKAKKSIAGFKLREGMPIGAKVTLRGERMYEFLDKLISVSLPRVRDFRGISKKA 121

Query: 123 FDGRGNITLGVKEQLIFPEINFDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK 180

FDGRGNITLGVKEQLIFPEI++D VDKVRG+D+VIVTTA+TDEE+RELL +GMPF K

45

Sbjct: 122 FDGRGNITLGVKEQLIFPEIDYDKVDRGMDVVIVTTASTDEEARELLSQMGMPFQK 179

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6357> which encodes the amino acid sequence <SEQ ID 6358>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1793(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

Identities = 177/180 (98%), Positives = 180/180 (99%)

-2323-

Query: 1 MANRLKEKYTNEVPALTEKFNYSSVMAVPKVEKIVLNMGVGDAVSNAKMLEKAAAEAL 60
 MANRLKEKYTNEV+PALTEKFNY+SVMAVPKVEKIVLNMGVGDAVSNAKMLEKAAAEAL
 Sbjct: 1 MANRLKEKYTNEVIPALTEKFNYTSVMAVPKVEKIVLNMGVGDAVSNAKMLEKAAAEAL 60

Query: 61 ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPT 120
 ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPT
 Sbjct: 61 ISGQKPLITKAKKSIAGFRLREGVAIGAKVTLRGERMYEFLDKLVSLSLPRVRDFHGVPT 120

Query: 121 KSFDGRGNYTLGVKEQLIFPEINFDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK 180
 KSFDGRGNYTLGVKEQLIFPEI+FDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK
 Sbjct: 121 KSFDGRGNYTLGVKEQLIFPEISFDDVDKVRGLDIVIVTTANTDEESRELLKGLGMPFAK 180

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2057

A DNA sequence (GBSx2169) was identified in *S.agalactiae* <SEQ ID 6359> which encodes the amino acid sequence <SEQ ID 6360>. This protein is predicted to be 50S ribosomal protein L24 (rplX). Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33285 GB:AF126061 Rpl24 [Streptococcus pneumoniae]
 Identities = 89/101 (88%), Positives = 94/101 (92%)

Query: 1 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKVVEGVALIKKHQKPNNENPQGAIVEKE 60
 MFVKKGDKVRVIAGKDKGTEAVVL ALPKVNKV+VEGV ++KKHQ+P NE PQG I+EKE
 Sbjct: 1 MFVKKGDKVRVIAGKDKGTEAVVLTALPKVNKVIVEGVNIVKKHQRPNTNLPQGGIIEKE 60

Query: 61 APIHVSNNVQVLDKNGVAGRVGYKVVDGKKVRYNKKSGEVL 101
 A IHVSNNVQVLDKNGVAGRVGYK VDGKKVRYNKKSGEVL
 Sbjct: 61 AAIHVSNNVQVLDKNGVAGRVGYKFVDGKKVRYNKKSGEVL 101

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6361> which encodes the amino acid sequence <SEQ ID 6362>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1850(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 95/101 (94%), Positives = 99/101 (97%)

Query: 1 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKVVEGVALIKKHQKPNNENPQGAIVEKE 60
 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKV+VEGV +IKKHQKPN ENPQGAIVEKE
 Sbjct: 1 MFVKKGDKVRVIAGKDKGTEAVVLKALPKVNKVIVEGVGMIKKHQKPNNTENPQGAIVEKE 60

Query: 61 APIHVSNNVQVLDKNGVAGRVGYKVVDGKKVRYNKKSGEVL 101
 APIHVSNNVQVLDKNGVAGR+GYKVVDGKKVRY+KKSGEVL

-2324-

Sbjct: 61 APIHVSINVQVLDKNGVAGRIGYKVVDGKKVRYSKKSGEVLD 101

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 2058

A DNA sequence (GBSx2170) was identified in *S.agalactiae* <SEQ ID 6363> which encodes the amino acid sequence <SEQ ID 6364>. This protein is predicted to be 50S ribosomal protein L14 (rplN). Analysis of this protein sequence reveals the following:

10 Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.1004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAD33284 GB:AF126061 RplL14 [Streptococcus pneumoniae]
 Identities = 116/122 (95%), Positives = 120/122 (98%)

Query: 1 MIQQETRLKVADNSGAREILTIKVLGGSGRKFFANIGDVIVASVKQATPGGAVKKGDVVKA 60
 MIQ ETRLKVADNSGAREILTIKVLGGSGRKFFANIGDVIVASVKQATPGGAVKKGDVVKA
 Sbjct: 1 MIQTETRLKVADNSGAREILTIKVLGGSGRKFFANIGDVIVASVKQATPGGAVKKGDVVKA 60

25 Query: 61 VIVRTKTGARRPDGSYIKFDDNAAVIRDDKTFRGTRIFGPVARELREGGYMKIVSLAPE 120
 VIVRTK+GARR DGSYIKFD+NAAVIR+DKTFRGTRIFGPVARELREGG+MKIVSLAPE
 Sbjct: 61 VIVRTKSGARRADGSYIKFDENAAVIRDDKTFRGTRIFGPVARELREGGFMKIVSLAPE 120

30 Query: 121 VL 122
 VL
 Sbjct: 121 VL 122

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6365> which encodes the amino acid sequence <SEQ ID 6366>. Analysis of this protein sequence reveals the following:

35 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.1004(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 122/122 (100%), Positives = 122/122 (100%)

Query: 1 MIQQETRLKVADNSGAREILTIKVLGGSGRKFFANIGDVIVASVKQATPGGAVKKGDVVKA 60
 MIQQETRLKVADNSGAREILTIKVLGGSGRKFFANIGDVIVASVKQATPGGAVKKGDVVKA
 Sbjct: 1 MIQQETRLKVADNSGAREILTIKVLGGSGRKFFANIGDVIVASVKQATPGGAVKKGDVVKA 60

50 Query: 61 VIVRTKTGARRPDGSYIKFDDNAAVIRDDKTFRGTRIFGPVARELREGGYMKIVSLAPE 120
 VIVRTKTGARRPDGSYIKFDDNAAVIRDDKTFRGTRIFGPVARELREGGYMKIVSLAPE
 Sbjct: 61 VIVRTKTGARRPDGSYIKFDDNAAVIRDDKTFRGTRIFGPVARELREGGYMKIVSLAPE 120

55 Query: 121 VL 122
 VL
 Sbjct: 121 VL 122

-2325-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2059

A DNA sequence (GBSx2171) was identified in *S.agalactiae* <SEQ ID 6367> which encodes the amino acid sequence <SEQ ID 6368>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3415(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAD33283 GB:AF126061 RpS17 [Streptococcus pneumoniae]
 Identities = 82/86 (95%), Positives = 83/86 (96%)

Query: 1 MERNQRKTLYGRVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60
 MERN RK L GRVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV
20 Sbjct: 1 MERNNRKVLVGRVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRPLSATKRFRIVEVVEKAVII 86
 RIMETRPLSATKRFRIVEVE+AVII
25 Sbjct: 61 RIMETRPLSATKRFRIVEVEEAVII 86

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6369> which encodes the amino acid sequence <SEQ ID 6370>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

30 bacterial cytoplasm --- Certainty=0.3415(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/86 (100%), Positives = 86/86 (100%)

40 Query: 1 MERNQRKTLYGRVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60
 MERNQRKTLYGRVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV
Sbjct: 1 MERNQRKTLYGRVSDKMDKTITVVVETKRNHPVYGKRINYSKKYKAHDENNVAKEGDIV 60

Query: 61 RIMETRPLSATKRFRIVEVVEKAVII 86
 RIMETRPLSATKRFRIVEVVEKAVII
45 Sbjct: 61 RIMETRPLSATKRFRIVEVVEKAVII 86

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2060

50 A DNA sequence (GBSx2172) was identified in *S.agalactiae* <SEQ ID 6371> which encodes the amino acid sequence <SEQ ID 6372>. Analysis of this protein sequence reveals the following:

Possible site: 43
>>> Seems to have no N-terminal signal sequence

-2326-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4329(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33282 GB:AF126061 RpL29 [Streptococcus pneumoniae]
 Identities = 58/68 (85%), Positives = 64/68 (93%)

10

Query: 1 MKLQEIKDFVKELRGLSQEELAKKENELKKELFDLRFQAAAGQLEKTARLDEVKKQIARV 60
 MKL E+K+FVKELRGLSQEELAK+ENELKKELF+LRFQAA GQLE+TARL EVKKQIAR+
 Sbjct: 1 MKLNEVKEFVKELRGLSQEELAKRENELKKELFELRFQAAATGQLEQTARLKEVKKQIARI 60

15

Query: 61 KTVQSEMK 68
 KTVQSE K
 Sbjct: 61 KTVQSEAK 68

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 **Example 2061**

A DNA sequence (GBSx2174) was identified in *S.agalactiae* <SEQ ID 6373> which encodes the amino acid sequence <SEQ ID 6374>. This protein is predicted to be RpL16 (rplP). Analysis of this protein sequence reveals the following:

Possible site: 52

25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4574(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33263 GB:AF126059 RpL16 [Streptococcus pneumoniae]
 Identities = 135/137 (98%), Positives = 137/137 (99%)

35

Query: 1 MLVPKRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
 MLVPKRVKHRREFRGKMRGEAKGGKEV+FGGEYGLQATTSHWITNRQIEAARIAMTRYMKR
 Sbjct: 1 MLVPKRVKHRREFRGKMRGEAKGGKEVAFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

40

Query: 61 GGVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEEVAREALRL 120
 GGVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEE+AREALRL
 Sbjct: 61 GGVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEEIAREALRL 120

45

Query: 121 ASHKLPVKCKFVKREAE 137
 ASHKLPVKCKFVKREAE
 Sbjct: 121 ASHKLPVKCKFVKREAE 137

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6375> which encodes the amino acid sequence <SEQ ID 6376>. Analysis of this protein sequence reveals the following:

50

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4574(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

-2327-

Identities = 136/137 (99%), Positives = 137/137 (99%)

Query: 1 MLVPRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60
 Sbjct: 1 MLVPRVKHRREFRGKMRGEAKGGKEVSFGEYGLQATTSHWITNRQIEAARIAMTRYMKR 60

Query: 61 GGVVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEEVAREALRL 120
 Sbjct: 61 GGVVWIKIFPHKSYTAKAIGVRMGSGKGAPEGWVAPVVRGKVMFEIAGVSEE+AREALRL 120

Query: 121 ASHKLPVKCKFVKREAE 137
 ASHKLPVKCKFVKREAE
 Sbjct: 121 ASHKLPVKCKFVKREAE 137

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2062

A DNA sequence (GBSx2175) was identified in *S. galactiae* <SEQ ID 6377> which encodes the amino acid sequence <SEQ ID 6378>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3758(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD33280 GB:AF126061 Rps3 [Streptococcus pneumoniae]
 Identities = 200/208 (96%), Positives = 203/208 (97%)

Query: 10 MRVGIIRDWDKAWAEKEYADYLHEDLAIRKFINKELADASVSTIEIERAVNKVIVSLHT 69
 MRVGIIRDWDKAWAEKEYADYLHEDLAIRKF+ KELADA+VSTIEIERAVNKV VSLHT
 Sbjct: 1 MRVGIIRDWDKAWAEKEYADYLHEDLAIRKFVQKELADAAVSTIEIERAVNKVNVSLHT 60

Query: 70 AKPGMVGKGGANVDALRGQLNKLTGKQVHINIIIEIKQPDLDHLVGENIARQLEQRVAF 129
 AKPGMVGKGGANVDALR +LNKLTGKQVHINIIIEIKQPDLDHLVGE IARQLEQRVAF
 Sbjct: 61 AKPGMVGKGGANVDALRAKLNKLTGKQVHINIIIEIKQPDLDHLVGEIARQLEQRVAF 120

Query: 130 RRAQKQAIQRTMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVP LHTLRADIDYAWEEAD 189
 RRAQKQAIQR MRAGAKGIKTQVSGRLNGADIARAEGYSEGTVP LHTLRADIDYAWEEAD
 Sbjct: 121 RRAQKQAIQRAMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVP LHTLRADIDYAWEEAD 180

Query: 190 TTYGKLGKVKVWIYRGEVLPARKNTKGGK 217
 TTYGKLGKVKVWIYRGEVLPARKNTKGGK
 Sbjct: 181 TTYGKLGKVKVWIYRGEVLPARKNTKGGK 208

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6379> which encodes the amino acid sequence <SEQ ID 6380>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3758(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2328-

Example 2063

A DNA sequence (GBSx2176) was identified in *S.agalactiae* <SEQ ID 6381> which encodes the amino acid sequence <SEQ ID 6382>. This protein is predicted to be 50S ribosomal protein L22 (rplV). Analysis of this protein sequence reveals the following:

5 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2704 (Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP: AAD33279 GB: AF126061 Rpl22 [Streptococcus pneumoniae]
 Identities = 99/114 (86%), Positives = 106/114 (92%)

 Query: 1 MAEITSAKAMARTVRVSPRKTRLVLDLIRGKNVADAIAILKFTPNKAARVIEKTLNSAIA 60
 MAEITSAKAMARTVRVSPRK+RLVLD IRGK+VADAIAIL FTPNKAA +I K LNSA+A
 Sbjct: 1 MAEITSAKAMARTVRVSPRKSRLVLDNIRGKSVADAIAILTFTPNKAAEIIKVLNSAVA 60

20 Query: 61 NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTHVTVVVSEK 114
 NAENNFGLEKANLVVSE FANEGPTMKRFRPRAKGSASPINKRT H+TV V+EK
 Sbjct: 61 NAENNFGLEKANLVVSEAFANEGPTMKRFRPRAKGSASPINKRTAHITVAVAEK 114

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6383> which encodes the amino acid sequence <SEQ ID 6384>. Analysis of this protein sequence reveals the following:

 Possible site: 36
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2794 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below.

 Identities = 113/114 (99%), Positives = 113/114 (99%)

 Query: 1 MAEITSAKAMARTVRVSPRKTRLVLDLIRGKNVADAIAILKFTPNKAARVIEKTLNSAIA 60
 MAEITSAKAMARTVRVSPRKTRLVLDLIRGK VADAIAILKFTPNKAARVIEKTLNSAIA
 40 Sbjct: 1 MAEITSAKAMARTVRVSPRKTRLVLDLIRGKKVADAIAILKFTPNKAARVIEKTLNSAIA 60

 Query: 61 NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTHVTVVVSEK 114
 NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTHVTVVVSEK
 Sbjct: 61 NAENNFGLEKANLVVSETFANEGPTMKRFRPRAKGSASPINKRTHVTVVVSEK 114

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2064

50 A DNA sequence (GBSx2177) was identified in *S.agalactiae* <SEQ ID 6385> which encodes the amino acid sequence <SEQ ID 6386>. This protein is predicted to be 30S ribosomal protein S19 (rpsS). Analysis of this protein sequence reveals the following:

 Possible site: 23
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

-2329-

bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 The protein is similar to ribosomal protein S19 from *S.pneumoniae*.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6387> which encodes the amino acid sequence <SEQ ID 6388>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 92/92 (100%), Positives = 92/92 (100%)

Query: 1 MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAYDGRKHVPV 60
 MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAYDGRKHVPV
 Sbjct: 19 MGRSLKKGPFVDEHLMKKVEAQANDEKKKVIKTWSRRSTIFPSFIGYTIAYDGRKHVPV 78

Query: 61 YIQEDMVGHLGGEFAPTRTYKGHAADDKKTRR 92
 YIQEDMVGHLGGEFAPTRTYKGHAADDKKTRR

Sbjct: 79 YIQEDMVGHLGGEFAPTRTYKGHAADDKKTRR 110

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2065

30 A DNA sequence (GBSx2178) was identified in *S.agalactiae* <SEQ ID 6389> which encodes the amino acid sequence <SEQ ID 6390>. This protein is predicted to be L2 (rplB). Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3182(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45959 GB:U43929 L2 [Bacillus subtilis]

Identities = 208/277 (75%), Positives = 239/277 (86%)

45 Query: 1 MGIVKYKPTTNGRRNMTSLDFAEIT'TNTPEKSLLVSLKNKAGRNNNGRITVRHQGGGHR 60
 M IK YKP++NGRR MT+ DFAEITT+ PEKSL L K GRNN G++TVRHQGGGHR
 Sbjct: 1 MAIKKYKPSSNGRRGMTTSDFAEIT'TDKPEKSL LAPLHKKGGRNNQGLTVRHQGGGHR 60

50 Query: 61 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYILAPKGLEVGQRIISGP 120
 YR+IDFKR+KDG+ V T+EYDPNR+ANIAL++Y DG K YILAPKG++VG ++SGP
 Sbjct: 61 QYRVIDFKRDKGIPGRVATVEYDPNRSANIALINYADGEKRYILAPKGIQVGTEVMMSGP 120

55 Query: 121 EADIKVGNALPLANIPVGTVIHNIELQPGKGAELIRAAGASAQVLGQEGKYVLVRLQSGE 180
 EADIKVGNALPL NIPVGTV+HNIEL+PGKG +L+R+AG SAQVLG+EGKYVLVRL SGE
 Sbjct: 121 EADIKVGNALPLINIPVGTVVHNIELKPGKGGLVRSAGTSAQVLGKEGKYVLVRLNSGE 180

Query: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRNRWKGVRPTVRGSMVNPNDHPHGGGEGKAP 240

-2330-

VRMIL CRA+IG VGNEQ L+NIGKAGR+RWKG+RPTVRGSMNPNDHPHGGGEG+AP
 Sbjct: 181 VRMILSACRASIGQVGNEQHELINIGKAGRSRWKGIRPTVRGSMNPNDHPHGGGEGRAP 240

Query: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRNQK 277
 +GRK+P +PWGKP LG KTR KK KSDK IVRRR K
 Sbjct: 241 IGRKSPSPWGPPLGFKTRKKKNSDKFIVRRRNK 277

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6391> which encodes the amino acid sequence <SEQ ID 6392>. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2560(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 264/277 (95%), Positives = 276/277 (99%)

Query: 1 MGIVKYKPTTNGRRNMTSLDFAEITTTNTPEKSLLVSLKKNKAGRNNGRITVRHQGGGHR 60
 +GIKVKYKPTTNGRRNMTSLDFAEITTT+TPEKSLLVSLK+KAGRNNGRITVRHQGGGHR
 Sbjct: 1 VGIVKYKPTTNGRRNMTSLDFAEITTTSTPEKSLLVSLKSKAGRNNGRITVRHQGGGHR 60

Query: 61 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYILAPKGLEVGQRIISGP 120
 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYI+APKGLEVGQRI+SGP
 Sbjct: 61 HYRLIDFKRNKDGVEAVVKTIEYDPNRTANIALVHYTDGVKAYIIAPKGLEVGQRIISGP 120

Query: 121 EADIKVGNALPLANIPVGTVIHNIELQPGKGAELIRAAGASAQVLGQEGKYVLVRLQSGE 180
 +ADIKVGNALPLANIPVGTV+HNIEL+PGKG EL+RAAGASAQVLGQEGKYVLVRLQSGE
 Sbjct: 121 DADIKVGNALPLANIPVGTVVHNIELKPGKGELVRAAGASAQVLGQEGKYVLVRLQSGE 180

Query: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRNKRWKGVPTVRGSMNPNDHPHGGGEGKAP 240
 VRMILGTCRATIGTVGNEQQSLVNIGKAGR+RWKG+RPTVRGSMNPNDHPHGGGEGKAP
 Sbjct: 181 VRMILGTCRATIGTVGNEQQSLVNIGKAGRSRWKGIRPTVRGSMNPNDHPHGGGEGKAP 240

Query: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRNQK 277
 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRN+K
 Sbjct: 241 VGRKAPSTPWGKPALGLKTRNKKAKSDKLIVRRRNEK 277

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2066

A DNA sequence (GBSx2180) was identified in *S.agalactiae* <SEQ ID 6393> which encodes the amino acid sequence <SEQ ID 6394>. This protein is predicted to be 50S ribosomal protein L23 (rplW). Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1669(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB03855 GB:AP001507 ribosomal protein L23 [Bacillus halodurans]
 Identities = 56/92 (60%), Positives = 67/92 (71%), Gaps = 1/92 (1%)

-2331-

Query: 2 NLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVTVK 61
 N DVIK+PVITE+S + KYTFEVD RA+K IK A+E FD VKVA VNT+ K
 Sbjct: 3 NARDVIKRPVITERSTEVMGDKKYTFEVDVRANKTQIKDAIEEIFD-VKVAKVNTMNYKG 61

5 Query: 62 KAKRVGRYTGFTSKTKKAIITLTADSKAIELF 93
 K KR GRYTGFT++ KKAI+TLT DSK ++ F
 Sbjct: 62 KPKRFGRYTGFTARRKKAIVTLTPDSKELDDFF 93

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6395> which encodes the amino acid
 10 sequence <SEQ ID 6396>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 96/98 (97%), Positives = 97/98 (98%)

Query: 1 MNLYDVIKKPVITEKSMVALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVTVK 60
 MNLYDVIKKPVITEKSM+ALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTV VK
 Sbjct: 1 MNLYDVIKKPVITEKSMIALEAGKYTFEVDTRAHKLLIKQAVEAAFDGVKVASVNTVNVK 60

25 Query: 61 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98
 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE
 Sbjct: 61 PKAKRVGRYTGFTSKTKKAIITLTADSKAIELFAAEAE 98

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 2067

A DNA sequence (GBSx2181) was identified in *S.agalactiae* <SEQ ID 6397> which encodes the amino
 acid sequence <SEQ ID 6398>. This protein is predicted to be 50S ribosomal protein L4 (rplD). Analysis of
 35 this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.54 Transmembrane 140 - 156 (139 - 156)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.1617(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45957 GB:U43929 L4 [Bacillus subtilis]
 Identities = 130/207 (62%), Positives = 160/207 (76%)

50 Query: 1 MANVKLFDQTGKEVSSVELNEAIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG 60
 M V L++Q G +ELN ++FGIEPNESVVFDD ++ QRASLRQGTH VKNRS V GG
 Sbjct: 1 MPKVALYNQNGSTAGDIELNASVFGIEPNESVVFDAIIMQRASLRQGTHKVKNRSEVRGG 60

Query: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSKVAE 120
 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSY YKLP+KVRRLA+KSV S+KV +
 55 Sbjct: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPKKVRRLAIAKSVLSSKVID 120

Query: 121 DKFVAVENLSFAAPKTAEFASVLSALSIDSKVLVILEEGNEFAALSARNLPNVTATATT 180
 + + +E+L+ KT E A++L LS++ K L++ + NE ALSARN+P VTV A

-2332-

Sbjct: 121 NNIIVLEDLTLDATAKTEMAAILKGLSVEKKALIVTADANEVALSARNIPGVTVEANG 180

Query: 181 ASVLDIVNADKLLVTKEAISTIEGVLA 207

+VLD+VN +KLL+TK A+ +E VLA

Sbjct: 181 INVLDVVNHEKLLITKAAVEKVEEVLA 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6399> which encodes the amino acid sequence <SEQ ID 6400>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2544(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 199/207 (96%), Positives = 203/207 (97%)

Query: 1 MANVKLFDQTGKEVSSVELNEAIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG 60

MANVKLFDQTGKEVSSVELN+AIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG

Sbjct: 1 MANVKLFDQTGKEVSSVELNDAIFGIEPNESVVFDDVISQRASLRQGTHAVKNRSVAVSGG 60

Query: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSKVAE 120

GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSKVAE

Sbjct: 61 GRKPWRQKGTGRARQGSIRSPQWRGGGVVFGPTPRSYGYKLPQKVRRLALKSVYSKVAE 120

Query: 121 DKFVAVENLSFAAPKTAEFASVLSALSIDS KVLVILEEGNEFAALSARNLPNVTVATATT 180

DKFVAVE LSFAAPKTAEFA VLSALSID+KVLV++EEGNEFAALSARNLPNVTVATA T

Sbjct: 121 DKFVAVEGLSFAAPKTAEFKVLALSIDS KVLVLEEGNEFAALSARNLPNVTVATAAT 180

Query: 181 ASVLDIVNADKLLVTKEAISTIEGVLA 207

ASVLDIVNADKLLVTKEAISTIE VLA

Sbjct: 181 ASVLDIVNADKLLVTKEAISTIEEVLA 207

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2068

A DNA sequence (GBSx2183) was identified in *S.galactiae* <SEQ ID 6401> which encodes the amino acid sequence <SEQ ID 6402>. This protein is predicted to be 50S ribosomal protein L3 (rplC). Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2090(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45956 GB:U43929 L3 [Bacillus subtilis]

Identities = 157/208 (75%), Positives = 180/208 (86%), Gaps = 2/208 (0%)

Query: 1 MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQKTVETDGYEAVQVGFDDKREVL 60

MTKGILG+K+GMTQ+F E+G+ IPVTVIEA PNVVLQ KT E DGYEA+Q+GFDDKRE L

Sbjct: 1 MTKGILGRKIGMTQVFAENGDLIPVTVIEAAPNVVLQKKTAEENDGYEAIQLGFDDKREKL 60

Query: 61 SNKPAKGHVAKANTAPKRFIREFKNIE--GLEVGAEELSVEQFEAGDVVDVTGTSGKGKFQ 118

-2333-

SNKP KGHVAKA TAPKRF++E + +E EVG E+ VE F AG++VDVTG SKGKGFQ
 Sbjct: 61 SNKPEKGHVAKAETAPKRFVKELRGVEMDAYEVGQEVKVEIFSAGEIVDVTGVS KGKGFQ 120
 Query: 119 GVIKRHGQSRGPMAGHSRYHRRPFGSMGPVAPNVRVFNKRLAGRMGGNRVTQVQNL EIVQVI 178
 5 G IKRHGQSRGPM+HGSRYHRRPFGSMGPV PNRVFK K L GRMGG ++TVQNL EIV+V
 Sbjct: 121 GAIKRHGQSRGPM SHGSRYHRRPFGSMGPVDPNVRVFKGLLPGRMGGEQITVQNL EIVKVD 180
 Query: 179 PEKNVVLKGNVPGAKKSLITIKSAVKA 206
 E+N++LIKGNVPGAKKSLIT+KSAVK+
 10 Sbjct: 181 AERNLLLIKGNVPGAKKSLITVKS AVKS 208

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6403> which encodes the amino acid sequence <SEQ ID 6404>. Analysis of this protein sequence reveals the following:

Possible site: 40
 15 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2090(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 205/208 (98%), Positives = 207/208 (98%)
 25 Query: 1 MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKT VETDGYEAVQVGFDDKREVL 60
 MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKT VETDGYEAVQVGFDDKREVL
 Sbjct: 1 MTKGILGKKVGMTQIFTESGEFIPVTVIEATPNVVLQVKT VETDGYEAVQVGFDDKREVL 60
 Query: 61 SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAE LSVEQFEAGDVVDVTGTSKGKGFQGV 120
 30 SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAE LSVEQFEAGDVVDVTG SKGKGFQGV
 Sbjct: 61 SNKPAKGHVAKANTAPKRFIREFKNIEGLEVGAE LSVEQFEAGDVVDVTGISKGKGFQGV 120
 Query: 121 IKRHGQSRGPMAGHSRYHRRPFGSMGPVAPNVRVFNKRLAGRMGGNRVTQVQNL EIVQVIPE 180
 IKRHGQSRGPMAGHSRYHRRPFGSMGPVAPNVRVFNKRLAGRMGGNRVTQVQNL EIVQVIPE
 35 Sbjct: 121 IKRHGQSRGPMAGHSRYHRRPFGSMGPVAPNVRVFNKRLAGRMGGNRVTQVQNL EIVQVIPE 180
 Query: 181 KNVVLKGNVPGAKKSLITIKSAVKA AK 208
 KNV+L+KGNVPGAKKSLITIKSAVKA AK
 40 Sbjct: 181 KNVILVKGNVPGAKKSLITIKSAVKA AK 208

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2069

A DNA sequence (GBSx2184) was identified in *S.agalactiae* <SEQ ID 6405> which encodes the amino
 45 acid sequence <SEQ ID 6406>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.43 Transmembrane 5 - 21 (5 - 21)
 50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2334-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2070

A DNA sequence (GBSx2185) was identified in *S.agalactiae* <SEQ ID 6407> which encodes the amino acid sequence <SEQ ID 6408>. This protein is predicted to be 30S ribosomal protein S10 (rpsJ). Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB46363 GB:L29637 S10 ribosomal protein [Streptococcus mutans]
Identities = 98/102 (96%), Positives = 102/102 (99%)

Query: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIIRATHKYKD 60
MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGA+VAGPVPLPTERSLYT+IRATHKYKD
Sbjct: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGASVAGPVPLPTERSLYTVIRATHKYKD 60
Query: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
SREQFEMRTHKRL+DI+NPTQKTVDALMKLDLPSGVNVEIKL
Sbjct: 61 SREQFEMRTHKRLIDIVNPTQKTVDALMKLDLPSGVNVEIKL 102

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6409> which encodes the amino acid sequence <SEQ ID 6410>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence
----- Final Results -----
bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 102/102 (100%), Positives = 102/102 (100%)
Query: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIIRATHKYKD 60
MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIIRATHKYKD
Sbjct: 1 MANKKIRIRLKAYEHRTLDTAAEKIVETATRTGATVAGPVPLPTERSLYTIIIRATHKYKD 60
Query: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102
SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL
Sbjct: 61 SREQFEMRTHKRLVDIINPTQKTVDALMKLDLPSGVNVEIKL 102

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2071

A DNA sequence (GBSx2186) was identified in *S.agalactiae* <SEQ ID 6411> which encodes the amino acid sequence <SEQ ID 6412>. Analysis of this protein sequence reveals the following:

Possible site: 34

-2335-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2538(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 2072

A DNA sequence (GBSx2187) was identified in *S.agalactiae* <SEQ ID 6413> which encodes the amino acid sequence <SEQ ID 6414>. Analysis of this protein sequence reveals the following:

15 Possible site: 46

>>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -11.41 Transmembrane 88 - 104 (79 - 110)
 INTEGRAL Likelihood = -8.39 Transmembrane 304 - 320 (300 - 324)
 INTEGRAL Likelihood = -6.58 Transmembrane 185 - 201 (180 - 206)
 INTEGRAL Likelihood = -5.63 Transmembrane 338 - 354 (331 - 357)
 INTEGRAL Likelihood = -5.52 Transmembrane 240 - 256 (237 - 259)
 INTEGRAL Likelihood = -4.99 Transmembrane 383 - 399 (375 - 407)
 INTEGRAL Likelihood = -3.82 Transmembrane 49 - 65 (48 - 73)
 INTEGRAL Likelihood = -2.87 Transmembrane 127 - 143 (121 - 144)
 INTEGRAL Likelihood = -2.81 Transmembrane 159 - 175 (159 - 177)
 INTEGRAL Likelihood = -2.18 Transmembrane 30 - 46 (30 - 47)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BAB06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 132/423 (31%), Positives = 210/423 (49%), Gaps = 16/423 (3%)

40 Query: 7 IIQLAIPAMIENTILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITTYQAIF--IALGASI 64
 + L P IE +L MLMG D +++Q AV+ V V+N I+ + +F +A G SI
 Sbjct: 11 L FALTWPIFIEILLHMLMGNADTLMLSQYSDDAVAAGVSNQILAVIIVMFGFVATGTSI 70

45 Query: 65 ASLLAKSLAGSKDDAISVCSQAIFLTLIGAVLGIISIVFGQTFKLLGTTKSVAVQVGG 124
 L+A+ L ++++A V +I L+ G VLG++ I FG K + S+ Q
 Sbjct: 71 --LVAQHLGAKERENAGKVAVVSIGANLIFGIVLGLLLIAFGPPILKAMQLDDSLQEQAT 128

50 Query: 125 LYLAIVGGGVVTLGMLTTLGSFLRVQGPRLPMYVSIFVNFLNAVLSGFAIFEWY----Y 180
 LYL IVGG V ++ T G+ LR + MYV+I +N LN + + IF
 Sbjct: 129 LYLQIVGGFSVQSLIMTAGAILRSHSFTKDVYMTIGMNIILNVIGNYLFIFGPFIPVPL 188

55 Query: 181 GLVGVAVSTLIARLIGICILAKYL-----PIKKIIKRMTWKISAQIWNLALPSAGER 232
 G+ GVA+ST+++R IG+ ++A L P ++KR + + +PSAGE+
 Sbjct: 189 GVTGVALSTVVSRTIGLFVIAILLYKRIRGELPFAYLLKRPFRVELRNLLKIGIPSAGEQ 248

 Query: 233 LMMRAGDVVIVAIVVQLGTNVVAGNAIGETLTQFNYPGLGIATATIILTAKYVGQKNRE 292
 L A +VI + +GT + + L F ++ + I T IL VG K +
 Sbjct: 249 LSYNASQLVITYFIAMMGTEALTTKVYTONLMMFVFLFAVAIGQGTQILIGHQVGAKQIQ 308

 Query: 293 SIEETIQSSYYIGLVLMILISSFMLLAGKPLTQLFTNPNPSAIGSLIVILLSFVGVGPATI 352
 + S +I + + + ++ PL +FT+NP + ++LL+ + P
 Sbjct: 309 AAYVRCFRSLWIAMTVSVSMVVFVFAFSTPLLGIFTDNPDLISLGTLLLLTILEPGRA 368

-2336-

Query: 353 GTLVYTAAWQGLGNAKLFPYTTTIGMWLIRVVLGYLLGIVFELGLLGVWMATIADNIFRW 412
 LV ++ + G+ K P Y + MW I V + YLLG+ LGL+GVW+A IAD FR
 Sbjct: 369 CNLVVISSLRAAGDVKFPVYLAIVSMWGIAPVPIAYLLGLPLGLGLIGVWIAFIADWFRG 428

Query: 413 LFL 415
 L +
 Sbjct: 429 LLM 431

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6415> which encodes the amino acid sequence <SEQ ID 6416>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -5.26 Transmembrane 89 - 105 (85 - 108)
 INTEGRAL Likelihood = -4.35 Transmembrane 305 - 321 (302 - 322)
 INTEGRAL Likelihood = -3.82 Transmembrane 161 - 177 (161 - 180)
 INTEGRAL Likelihood = -3.82 Transmembrane 192 - 208 (189 - 208)
 INTEGRAL Likelihood = -3.77 Transmembrane 129 - 145 (128 - 151)
 20 INTEGRAL Likelihood = -3.24 Transmembrane 242 - 258 (240 - 258)
 INTEGRAL Likelihood = -2.81 Transmembrane 378 - 394 (377 - 394)
 INTEGRAL Likelihood = -2.66 Transmembrane 339 - 355 (338 - 358)
 INTEGRAL Likelihood = -2.60 Transmembrane 58 - 74 (58 - 75)
 INTEGRAL Likelihood = -2.50 Transmembrane 32 - 48 (32 - 49)

25 ----- Final Results -----

bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 The protein has homology with the following sequences in the databases:

>GP:BAB06655 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 119/435 (27%), Positives = 214/435 (48%), Gaps = 14/435 (3%)

35 Query: 9 IFSLALPSMIENILQMLMGMVDNYLVAQIGLVAVSGVSIANNIISIYQSLFIALGAAVSS 68
 +F+L P IE +L MLMG D +++Q AV+ V ++N I+++ +F + S
 Sbjct: 11 LFALTWPFIETIILLHMLMGNADTLMLSQYSDDAVAAGVSNQILAVIIVMFGFVATGTISI 70

40 Query: 69 LIARSIGENNQNKQLNVMAGVLQVTLTLLSVGLGLLSVAGHHQVLEWLGAESVTLVGGQY 128
 L+A+ +G + + L+ + LGLL +A +L+ + + S+ Y
 Sbjct: 71 LVAQHLGAKERENAGKVAVVSIGANLIFGIVLGLLLIAFGPPILKAMQLDDSLLEATLY 130

45 Query: 129 LSIVGGMIVSLGLTSLGAIVRAQGYPKIPMQVSLINVLNIAIFSALSIY----VWGFGL 184
 L IVGG V L+ + GAI+R+ + K M V++ +N+LN I + L I+ + G+
 Sbjct: 131 LQIVGGFSVVQSLIMTAGAILRSHSFTKDVMTYIGMNILNVIGNYLFIFGPPGIPVLGV 190

50 Query: 185 LGVAWATVLSRLVGVFLLCQF-----IPIKQVAKRLMRPLDKIIFDLSLPAAGERLM 236
 GVA +TV+SR +G+F++ +P + KR R + + + +P+AGE+L
 Sbjct: 191 TGVALSTVVSRTIGLFVIAILLYKRIRGELPFAYLLKRFPVELRNLLKIGIPSAGEQLS 250

55 Query: 237 MRAGDVLIIGIVVRFGTALAGNAIGETLTQFNYPGLAMATATIILVARQLGGGKVTEI 296
 A ++I + GT AL + L F ++ +A+ T IL+ Q+G ++
 Sbjct: 251 YNASQLVITYFIAMMGTEALTTKVYTQNLMMFVFLFAVAIGQGTQILIGHQVGAKQIQAA 310

Query: 297 RYIIREAFILSTLMLVMGALTYLLGPSLLPLFTQNTDAQRSAMIVLLFSLLGAPATAGT 356
 + ++ + + M + + LL +FT N D +LL +++ P A
 Sbjct: 311 YVRCFRSLWIAMTVSVSMVAVVFAFSTPLLGIPTDNPDIISLGTTLTLLLTIIIEPGRACN 370

60 Query: 357 LVYTAVWQGLGKAKLPFYATTIGMWVIRIGLGYVIGVWQYGLIGVWMATVLDNTRWFI 416
 LV + + G K P Y + MW I + + Y++G+ GLIGVW+A + D R +
 Sbjct: 371 LVVISSLRAAGDVKFPVYLAIVSMWGIAPVPIAYLLGLPLGLGLIGVWIAFIADWFRGLL 430

65 Query: 417 LSKHFK--KYQETIF 429
 + ++ K+QE++F
 Sbjct: 431 MIWRWRKKGKQEMSF 445

-2337-

An alignment of the GAS and GBS proteins is shown below.

Identities = 219/418 (52%), Positives = 316/418 (75%)

```

5  Query: 5  KEIIQLAIPAMIENTILQMLMGVVDNYLVAQLGVVAVSGVSVANNIITIYQAI FIALGASI 64
    ++I LA+P+MIENILQMLMG+VDNYLVAQ+G+VAVSGVS+ANNII+IYQ++FIALGA++
    Sbjct: 7  RKIFSLALPSMIENILQMLMGVVDNYLVAQIGLVAVSGVSIANNIISIYQSLFIALGAAV 66

    Query: 65  ASLLAKSLAGSKKDDAISVCSQAIFLTLLIGAVLGIISIVFGQTFKLLGTTKSVAQVGG 124
    +SL+A+S+ + ++ ++ + + +TLL+ LG++S+ + LG SV VGG
10  Sbjct: 67  SSLIARSIGENNQNKLQNYMAGVLQVTLTLLSVGLGLLSVAGHHQVLEWLGAESVTLVGG 126

    Query: 125  LYLAIVGGGVVTLGMLTTLGSLRVQGPRLPMYVSIFVNFNLAIVLGGFAIFEWRYGLVG 184
    YL+IVGG +V+LG+LT+LG+ +R QG P++PM VS+ +N LNA+ S +I+ W +GL+G
15  Sbjct: 127  QYLSIVGGMIVSLGLLTSGLAIVRAQGYPKIPMQVSLINVLNAIFSALSIYVWGFGLLG 186

    Query: 185  VAVSTLIARLIGICILAKYLPKIKIKRMTWKISAQIWNALPSAGERLMMRAGDVVIVA 244
    VA +T+++RL+G+ +L +++PIK++ KR+ + I++L+LP+AGERLMMRAGDV+I+
    Sbjct: 187  VAWATVLSRLVGVFLLCQFPIPIKQVAKRLMRPLDKIIFDLSLPAAGERLMMRAGDVLIIG 246

20  Query: 245  IVVQLGTNVVAGNAIGETLTQFNYPGLGIATATIILTAKYVGQKNRESIEETIQSSYYI 304
    IVV+ GT +AGNAIGETLTQFNYPGL +ATATIIL A+ +G I I+ ++ +
    Sbjct: 247  IVVRFGTTALAGNAIGETLTQFNYPGLAMATATIILVARQLGGGKVTEIRYIIREAFIL 306

    Query: 305  GLVLMILISSFMLLAGKPLTQLFTNNPSAIKGSILIVILLSFVGVPATIGTLVYTAAWQGL 364
    ++M+++ + L G L LFT N A + ++IV+L S +G PAT GTLVYTA WQGL
25  Sbjct: 307  STLMLVMGALTYLLGPSLLPLFTQNTDAQRSAMIVLLFSLGAPATAGTLVYTAVWQGL 366

    Query: 365  GNAKLPHYTTTIGMWLIRVVLGYLLGIVFELGGLGVWMATIADNIFRWLFLKVHYHRY 422
    G AKLPFY TTIGMW+IR+ LGY++G+V++ GL+GVWMAT+ DN RW L H+ +Y
30  Sbjct: 367  GKAKLPFYATTIGMWVIRIGLGYVIGVWVWQYGLIGVWMATVLDNTSRWFILSKHFKKY 424
    Identities = 48/211 (22%), Positives = 89/211 (41%), Gaps = 29/211 (13%)

    Query: 213  MTWKISAQIWNALPSAGERLMMRAGDVVIVAIVVQLGTNVVAGNAIGETLTQFNYPGL 272
    M + +I++LALPS E ++ +V +V Q+G V+G +I + +
35  Sbjct: 1  MIYNNRRKIFSLALPSMIENILQMLMGVVDNYLVAQIGLVAVSGVSIANNIISIYQSLFI 60

    Query: 273  GIATATIILTAKYVGQKNRESIEETIQSSYYIGLVLMILISSFML-----L 318
    + A L A+ +G+ N+ Q +Y G++ + L+ S L L
40  Sbjct: 61  ALGAAVSSLIARSIGENNQNK-----QLNYMAGVLQVTLTLLSVGLGLLSVAGHHQVLEWL 115

    Query: 319  AGKPLTQLFTNNPSAIKGSILIVILLSFVGVPATIGTLVYTAAWQGLGNAKLPHYTTTIGM 378
    + L +I G +IV L G+ ++G +V + G K+P + + +
    Sbjct: 116  GAESVTLVGGQYLSIVGGMIVSL----GLLTSGLAIV----RAQGYPKIPMQVSLI-I 165

45  Query: 379  WLIRVVLGYLLGIVFELGGLGVWMATIADNI 409
    ++ + L V+ GLLGV AT+ +
    Sbjct: 166  NVLNAIFSALSIYVWGFGLLGVAWATVLSRL 196

```

A related GBS gene (SEQ ID 8071) and protein (SEQ ID 8072) were also identified. Analysis of this

```

INTEGRAL    Likelihood = -2.18   Transmembrane   30 - 46 ( 30 - 47)
PERIPHERAL  Likelihood =  0.32      11
modified ALOM score:   2.78

```

~~~~~ Final Results ~~~~~

The protein has homology with the following sequences in the databases:

48                    78                    108                    138                    168                    198                    228                    258  
YK\*RRDTGFRCYFNLKRFVRCFFT\*GGYRSTKGRSNP\*NGSTYLKYARNG\*RVSRFETIIKIRLF\*NI\*SEKETF\*KFSH

288            318            348            378            408            438            468            498  
HSLFNDPG\*\*KGD TVRYSKEI IQLAIPAMIE NILQMLMGVVDNYLVAQLGVAVSGVSVANNIIIT YQAIFIALGASIAS  
||::|::|| |::| ||| :| ::| :|::| :| :| :|  
RYSLFKNYLPKEEVPEIRKELIKLALPAMGENVLQMLFGMADTAF LGHSYWKAMS GVGLSNQVFVVVQVLIAASMGATV  
20                 30                 40                 50                 60                 70                 80

759            789            819            837            867            897            909            939  
GSFLRVQGQPRLPYVSIFVNFNLAVLSGFAIF---EWRYGLVGVAVSTLIARLIGICILA-----KYLPIKKIIKRM  
: || | || | : : ||| | || | : | || : : : : || : : : :  
SAMLRGAGDTRTPMIVTGLTNFLNIFLDYAMIFGKFGFPEMGVRGAAVATILSRFVGAGILTYVIFKREEFQLRKGLVPP  
170            180            190            200            210            220            230

1209            1239            1269            1299            1329            1359            1383            1413  
IEETIQSSYYIGLVLMILISSFMLLAGKPLTQLFTNPSAIGSLIVILLSSFVGVGPATIGTLVYT--AAWQGLGNAKLFF  
:     |:   :   :   ::   :   :   ::|   :||   ::||::|   |:   |   :   :   :|:   :   |   |   :|   ||   |  
VLGVIRQGWLISLLFQVTVGIIIFLFPEPLIRIFTSDPQIIEISKLPV--KIIGLFQFFLAIDSTMNGALRGNTGNTLPPM  
  
              330            340            350            360            370            380            390

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2339-

**Example 2073**

A DNA sequence (GBSx2188) was identified in *S.agalactiae* <SEQ ID 6417> which encodes the amino acid sequence <SEQ ID 6418>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2200(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD05671 GB:AE001448 THREONINE SYNTHASE [Helicobacter pylori  
J99]

Identities = 161/479 (33%), Positives = 259/479 (53%), Gaps = 17/479 (3%)

Query: 14 KVTASQAILKGLADDGGLFTPTITFPKVDLDFTKLKDASYQEVAKLVLSAFFDDFTEQELD 73  
K+ +A+L A GGL+T F L++ SY E+ + V + + L

Sbjct: 13 KIDFIEAVLNPNAKPGGLYTLLEHFET--LEWQDCLGMSYSELVEHVFEELLNLEIPKNLLA 70

Query: 74 YCISQAYDTKFDTEIAPIVKIGDRYHL-ELFHGPTIAFKDMALSILPYLLTTAAKKQGV 132  
+ + Y+ + API + +R + EL+HGP++AFKDMAL L L + A G

Sbjct: 71 SALKR-YENFDNPKNPAPIFALNERLFVQELYHGPSLAFKDMALQPLASLFSNLAV--GK 127

Query: 133 DNKIVILTATSGDTGKAAMAGFADVPGETEIVFYPKNGVSYIQELQMITQAGQNTHVVAI 192  
+ K ++L +TSGDTG A + G A +P ++ YPK+G S +Q+LQM+TQ N V +

Sbjct: 128 NEKYLVLVSTSGDTGTPATLEGLAGMPNVFVCLYPKDGTSLVQKLMQVTQNASNLKVFVG 187

Query: 193 EGNFDDAQTSVKEMFNNSLLRLKLSQQHMQLSANSNMNIGRLVLPQIVYYIYAYAQLVKSK 252  
G+FDDAQ ++K + + L + ++LS ANS+N GR+ QIVY+I+ + +L K

Sbjct: 188 SGDFDDAQNALKNLLKDDDFNEALKARQLKLSVANSVNFGRIFQIVYHIWGFLELYKKG 247

Query: 253 EISIGQPINFVSPTGNFGNLAAYASQIGLPTVKLICASNDNNVLTDFFKTQTYD-KNR 311  
I+ + I ++P+GNFGN L A+YA ++GL + K+ +N N+VL +F +T YD R

Sbjct: 248 AINSKEKITLAIPSGNFGNALGAFYAKMGLNIAKIKVVTNSNDVLEFIEETGRYDLTKR 307

Query: 312 EFKVTSSPSMDILVSSNLERLIFHLLGDDAETTKLMEDLVTTGEYALEARQANIL-ESF 370  
K T SP+MDIL SSN+ER +F L G E T +LM+ L YAL+ ++ +L E F

Sbjct: 308 SLKQTFSPAMDILKSSNVERALFSLFG--FERTLELMQALEEEKFYALKPKELALLQEHF 365

Query: 371 VAGFATEQFVELDIKHLFDQYQYIEDPHTAVASAVYQAYQTETKDQTPAVIVSTASPYKF 430  
+++ I+ ++ ++QY+ DPHTA A K ++ +TAS KF

Sbjct: 366 SCASCSDEDCLKTIQEVYAEHQYLIDPHTAT-----ALNASLKTHEKTLVSATASYEKF 419

Query: 431 PCVVTKAIT-NKEEIQDFAAISILNDLSGVSLPKAVITDLQKAEVIHRTVVPTSNMRETV 488  
P A+ K+ D AA+ L + + + DL + + H+ V+ + + + +

Sbjct: 420 PKTLLALNEQKKNDKAALETLLKNSYNTPDSQRLLDLFERGIKHQEVLEKLNEIKSSI 478

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2074**

A DNA sequence (GBSx2189) was identified in *S.agalactiae* <SEQ ID 6419> which encodes the amino acid sequence <SEQ ID 6420>. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2340-

bacterial cytoplasm --- Certainty=0.3153(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9279> which encodes amino acid sequence <SEQ ID 9280> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF40975 GB:AE002410 alcohol dehydrogenase, propanol-preferring  
 [Neisseria meningitidis MC58]  
 10 Identities = 202/282 (71%), Positives = 228/282 (80%), Gaps = 1/282 (0%)

Query: 1 MGHEGIGIVEEIGEGVTSRLVGDRVSIWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 60  
 +GHEGIG+V+E+ +GV +L+VGDRVSIW F+ CG CEYC TGRETLCRSV NAGY+ DG  
 15 Sbjct: 60 LGHEGIGLVKEVADGVKNLKVGDVSIWLFQSCGSCEYCNTGRETLCRSVLNAGYTADG 119

Query: 61 GMSEYAIVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIAVYGAGGLGN 120  
 GM+ + IV+ADYAVKVPEGLDPAQASSITCAGVTTYKAIK +G PGQWIA+YGAGGLGN  
 Sbjct: 120 GMATHCIVSADYAVKVPEGLDPAQASSITCAGVTTYKAIKVSQVRPGQWIAIYGAGGLGN 179

20 Query: 121 LAVQYAKKVFNAHVVAVDINADKLQLAKEVGADLTVNGKEIKDVAAYIQEKTGGCHGVVV 180  
 L VQYAKKVF AHVVA+DIN DKL AKE GADL VN + +D A IQEKTGG H VV  
 Sbjct: 180 LGVQYAKKVFNAHVVAIDINDDKLAFAKETGADLVVNAK-EDAAKVIQEKTGGHAHAHV 238

25 Query: 181 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGIRVVGSLVGTTRKDLEEF 240  
 TAVS AFN A++ VRAGG VVA+GLP E M+LSI + VLDGI VVGSVGTTRKDLEEF  
 Sbjct: 239 TAVSAAAFNSAVNVCVRAGGRVVAIGLPPESMDLSIPRLVLDGIEVVGSVGTTRKDLEEF 298

Query: 241 AFGAEGLVVPVVEKVPVDTAPQVFDEMERGLIQGRKVLDF 282  
 FGAEGLVVP V+ +D AP +F EM G I GR V+D K  
 30 Sbjct: 299 QFGAEGLVVPKVQLRALDEAPAIQEMREGKITGRMVIDMKK 340

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6421> which encodes the amino acid sequence <SEQ ID 6422>. Analysis of this protein sequence reveals the following:

35 Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2356(Affirmative) < succ>  
 40 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 263/280 (93%), Positives = 273/280 (96%)

45 Query: 1 MGHEGIGIVEEIGEGVTSRLVGDRVSIWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 60  
 +GHEGIGIVEEIGEGVTSRL+VGDRVSIWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG  
 Sbjct: 76 LGHEGIGIVEEIGEGVTSRLVGDRVSIWFFEGCGHCEYCTTGRETLCRSVKNAGYSVDG 135

50 Query: 61 GMSEYAIVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIAVYGAGGLGN 120  
 GMSEYA+VTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWI ++GAGGLGN  
 Sbjct: 136 GMSEYAVVTADYAVKVPEGLDPAQASSITCAGVTTYKAIKEAGAAPGQWIVIFGAGGLGN 195

55 Query: 121 LAVQYAKKVFNAHVVAVDINADKLQLAKEVGADLTVNGKEIKDVAAYIQEKTGGCHGVVV 180  
 LAVQYAKKVFNAHVVAVDIN DKL+LAKEVGAD+ VNGKEI+DV YIQEKTGG HGVVV  
 Sbjct: 196 LAVQYAKKVFNAHVVAVDINDKLELAKEVGADILVNGKEIEDVPGYIQEKTGGAGHVVV 255

Query: 181 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGIRVVGSLVGTTRKDLEEF 240  
 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGI+VVGSLVGTTRKDLEEF  
 Sbjct: 256 TAVSKVAFNQAIDSVRAGGTVVAVGLPSEYMELSIVKTVLDGIKVVGSVGTTRKDLEEF 315

60 Query: 241 AFGAEGLVVPVVEKVPVDTAPQVFDEMERGLIQGRKVLDF 280  
 AFGAEGLV VPVEKVPVDTAP+VFDEMERGLIQGRKVLDF

-2341-

Sbjct: 316 AFGAEGLVAPVVEKVPVDTAPEVFDEMERGLIQGRKVLDF 355

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 5 Example 2075

A DNA sequence (GBSx2190) was identified in *S.agalactiae* <SEQ ID 6423> which encodes the amino acid sequence <SEQ ID 6424>. Analysis of this protein sequence reveals the following:

Possible site: 42

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

|    |          |                    |               |                        |
|----|----------|--------------------|---------------|------------------------|
| 10 | INTEGRAL | Likelihood = -9.82 | Transmembrane | 83 - 99 ( 76 - 108)    |
|    | INTEGRAL | Likelihood = -7.27 | Transmembrane | 46 - 62 ( 43 - 65)     |
|    | INTEGRAL | Likelihood = -7.22 | Transmembrane | 187 - 203 ( 182 - 209) |
|    | INTEGRAL | Likelihood = -6.00 | Transmembrane | 243 - 259 ( 229 - 262) |
|    | INTEGRAL | Likelihood = -4.25 | Transmembrane | 404 - 420 ( 402 - 422) |
| 15 | INTEGRAL | Likelihood = -3.98 | Transmembrane | 120 - 136 ( 119 - 136) |
|    | INTEGRAL | Likelihood = -3.88 | Transmembrane | 308 - 324 ( 307 - 324) |
|    | INTEGRAL | Likelihood = -2.13 | Transmembrane | 378 - 394 ( 376 - 394) |
|    | INTEGRAL | Likelihood = -1.38 | Transmembrane | 152 - 168 ( 152 - 168) |
| 20 | INTEGRAL | Likelihood = -1.17 | Transmembrane | 271 - 287 ( 271 - 287) |

----- Final Results -----

bacterial membrane --- Certainty=0.4927(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9371> which encodes amino acid sequence <SEQ ID 9372> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:AAC17857 GB:AF026147 Yoji [Bacillus subtilis]

Identities = 183/432 (42%), Positives = 266/432 (61%), Gaps = 1/432 (0%)

Query: 1 MKLFIPVLIYQFANFSATFIDSVMTGQYSQHLHLAGVSTASNLWTPPFALLVGMISALVPV 60

+ + IP+ I Q TF+D+VM+G+ S LAGV+ S+LWTP + L G++ A+ P+

Sbjct: 15 LHILIPFITQAGLSLITFLDTVMGKVPSPADLAGVAIGSSSLWTPVVTGLAGILMAVTPI 74

Query: 61 VGQHLGRGNKEQIRTEFHQFLYLGLILSLILFLIMQFIAQPVLSLGLLEDEVAVGRGYL 120

V Q LG K++I Q +Y+ +LS+ + +I +LG L L+ V + + +L

Sbjct: 75 VAQLLGAEEKKQKIPFTVLQAVVVAALLSIAVLVIGYAAVDLILGRLNLDIHVHQIAKHFL 134

Query: 121 NYMLIGIMPLVLFSICRSFFDALGLTRLNMYLMLLILPFNSFFNYMLIYKFGMPRLGGA 180

++ +GI PL ++++ RSF D+LG TR++M + L LP N NY+ I+GKFGMP LGG

Sbjct: 135 GFLSLGIPPLFVYTVLRSFIDSLGKTRVTMMITLSSLPINFVLNVVFIFGKFGMPALGGV 194

Query: 181 GAGLGTSLTYWAIFIVIIIVMSLHPQIKTYHIW-TLERIKAPLIIEDIRLGLPIGLQIFA 239

GAGL ++LTYW I I+ ++ + Y I+ T+ + +++GLPIG +F

Sbjct: 195 GAGLASALTYWCICIIISFFIIHKNAPFSEYGIFLTMYKFSWKACKNLLKIGLPIGFAVFF 254

Query: 240 EVAIFAVVGLFMAKFSSIIIAAHQAAMNFSSLMYAFPLSISTALAITISFEVGAERFQDA 299

E +IFA V L M+ F ++ IA+HQAAMNF+SL+Y PLS+S AL I + FE GA RF+DA

Sbjct: 255 ETSIFAAVTLLMSHFHTVTIASHQAAMNFASLLYMLPLSVSMALTIVVGFEAGAARFKDA 314

Query: 300 NTYSRIGRLTAVGITSGLTLLFLFLFRENVAAMYNSDPHFVAITAQFLTYSLFFQFADAYA 359

+YS IG + A+G + T + LFRE +A MY SDP + +T FL Y+LFFQ +DA A

Sbjct: 315 RSYSLIGIMMAIGFSLFTAACILLFREQIAGMYTSDPDVRLTQHFLLIYALFFQLSDAVA 374

Query: 360 APVQGILRGYKDTTKPFMIGAGSYWLCALPLAVILEKNSQLGPFAYWIGLITGIFVCGLF 419

AP+QG LRGYKD SYW+ LP+ ++ + LG F YWIGLI G+ +

Sbjct: 375 APIQGALRGYKDVNYTLAAAFVSYWVIGLPVGYMVGTFSTLGAFGYWIGLIAGLAAGAVG 434

Query: 420 LNQRLOKIKKLY 431

-2342-

L RL K++K Y  
 Sbjct: 435 LFFRLAKLQKRY 446

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2076

A DNA sequence (GBSx2191) was identified in *S.agalactiae* <SEQ ID 6425> which encodes the amino acid sequence <SEQ ID 6426>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.60 Transmembrane 23 - 39 ( 23 - 39)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2077

A DNA sequence (GBSx2192) was identified in *S.agalactiae* <SEQ ID 6427> which encodes the amino acid sequence <SEQ ID 6428>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3829(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC06891 GB:AE000703 hypothetical protein [Aquifex aeolicus]  
 Identities = 72/213 (33%), Positives = 115/213 (53%), Gaps = 11/213 (5%)  
 Query: 36 RPKILMHVCCAPCSTYTLEYLSQ---WADVTIYFANSNIHPKDEYYRREYVTQKFVHDFN 92  
 + KIL+H+CCAP + Y L+ L + +++ YF + NIHP +EY R T++ +  
 Sbjct: 3 KSKILVHICCAPDAIYFLKKLREDYPESEIIIGYFYDENIHPYBEYRLRYLETERICKELG 62  
 Query: 93 KNTGYSVQFLSAPYEPNEFFKIVHGLEEEPEGDRCKVCYDFRLDKTAEKAVELGFDFYFG 152  
 N + Y+ + + V G E+EPE G RC++C+D+RL+K+AE A ELG D  
 Sbjct: 63 IN-----LIEGEYDLENWLERVKGYEDEPERGKRCQICFDYRLEKSAEVAKELGCDALT 116  
 Query: 153 SALTISPHKNSQTINTIGIDVQKIYDTQYLPSDLKKNKGYSVEMCKDYDIYRQCYCGC 212  
 + L +SP K+ + G + K ++L D +K G Q ++ K+ +IY+Q YCGC  
 Sbjct: 117 TTLMSPPKKSIPQLKKAGEEATKRTGIEFLAPDYRKGGGTQEMFKLSKEREIYQQDYCGC 176  
 Query: 213 IFGAKDQGINLLQIKKDAKAFVSDKDGKEEFPN 245  
 I+G Q +I D F+ + G +E N  
 Sbjct: 177 IYGLFKQKNG--KIFWDLVGFLGRRPGSKEERN 207



-2343-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6429> which encodes the amino acid sequence <SEQ ID 6430>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.3498(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 RGD motif: 254-256

The protein has homology with the following sequences in the databases:

>GP:AAC06891 GB:AE000703 hypothetical protein [Aquifex aeolicus]  
 Identities = 65/182 (35%), Positives = 106/182 (57%), Gaps = 9/182 (4%)  
 Query: 39 RPSILMHVCCAPCSTYTLEYLTQF---ADITVYFANSNIHPKDEYHRRAYVTQQFVSEFN 95  
           + IL+H+CCAP + Y L+ L + ++I YF + NIHP +EY R T++ E  
 Sbjct: 3 KSKILVHICCAPDAIYFLKKLREDYPESEIIGYFYDPNIHPYEEYRLRYLETERICKELG 62  
 Query: 96 AKTGNTVQFLEADYVPNEYVRQVRGLEEPEGGDRVCFCFDYRLDKTAQKAVELGFDYFA 155  
           + +E +Y ++ +V+G E+EPE G RC++CFDYRL+K+A+ A ELG D  
 Sbjct: 63 -----INLIEGEYDLENWLERVKGYEDEPERGKRCQICFDYRLKSAEVAKELGCDALT 116  
 Query: 156 SALTISPHKNSQTINDVGIDVQKVYTTKYLPSDFKKNNGYRRSVEMCEEYDIYRQCYCGC 215  
           + L +SP K+ + G + K ++L D++K G + ++ +E +IY+Q YCGC  
 Sbjct: 117 TTLMSPPKKSIPQLKKAGEEATKRTGIEFLAPDYRKGGGTQEMFKLSKEREIYQQDYCGC 176  
 Query: 216 VY 217  
           +Y  
 Sbjct: 177 IY 178

An alignment of the GAS and GBS proteins is shown below.

Identities = 184/255 (72%), Positives = 219/255 (85%)  
 Query: 1 MIDVENILEKMKPNQKINYDWMQOMVKQWQASDIRPKILMHVCCAPCSTYTLEYLSQWA 60  
           MID++ IL M PNQKINYD VMQQM K W+ +RP ILMHVCCAPCSTYTLEYL+Q+A  
 Sbjct: 4 MIDLQEILANMNPQKINYDRVMQMAKVWEKESVRPSILMHVCCAPCSTYTLEYLTQFA 63  
 Query: 61 DVTIYFANSNIHPKDEYRREYVTQKFVHDFNKNTGYSVQFLSAPYEPNEFFKIVHGLEE 120  
           D+T+YFANSNIHPKDEY+RR YVTQ+FV +FN TG +VQFL A Y PNE+ + V GLEE  
 Sbjct: 64 DITVYFANSNIHPKDEYHRRAYVTQQFVSEFNKNTGNTVQFLEADYVPNEYVRQVRGLEE 123  
 Query: 121 EPEGGDRCKVCYDFRLDKTAEKAVELGFDYFGSALTISPHKNSQTINTIGIDVQKIYDTQ 180  
           EPEGGDRC+VC+D+RLDKTA+KAVELGFDYF SALTISPHKNSQTIN +GIDVQK+Y T+  
 Sbjct: 124 EPEGGDRVCFCFDYRLDKTAQKAVELGFDYFASALTISPHKNSQTINDVGIDVQKVYTTK 183  
 Query: 181 YLPSDLKKNKGYSVEMCKDYDIYRQCYCGCIFGAKDQGINLLQIKDAKAFVSDKDGD 240  
           YLPSD KKN GY+RSVEMC++YDIYRQCYCGC++ AK QGI+L+Q+KKDAKAF++DKD  
 Sbjct: 184 YLPSDFKKNNGYRRSVEMCEEYDIYRQCYCGCVYAAKMGIDLVQVKDAKAFMADKDL 243  
 Query: 241 EEFPNIRFTFNGKSM 255  
           +F +IRF++ G M  
 Sbjct: 244 NDFTHIRFSYRGDEM 258

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2078

A DNA sequence (GBSx2193) was identified in *S.agalactiae* <SEQ ID 6431> which encodes the amino acid sequence <SEQ ID 6432>. Analysis of this protein sequence reveals the following:

-2344-

Possible site: 53

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4216(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB14809 GB:Z99118 excinuclease ABC (subunit C) [Bacillus subtilis]  
 Identities = 189/333 (56%), Positives = 244/333 (72%)

Query: 1 MNELIKHKLELLPDSPGCGYLHKDKNGTIIYVGKAKNLKNRVKSYFHGSHNTKTELLVSEI 60  
 MN+ +K KL LLPD PGCYL KD+ T+IYVGKAK LKNRV+SYF GSH+ KT+ LV+EI  
 15 Sbjct: 1 MNKQLKEKLALLPDQPGCYLMKDRQQTVIYVGKAKVLKNRVRSYFTGSHDAKTQRLVTEI 60

Query: 61 EDFEYIVTTSNTEALLLEINLIQENMPKYNIRLKDDKSYPYIKITNERYPRLMITRQVKK 120  
 EDFEYIVT+SN EAL+LE+NLI+++ PKYN+ LKDDK+YP+IK+T+ER+PRL++TR VKK  
 Sbjct: 61 EDFEYIVTSSNLEALILEMNLIKKHDPKYNVMLKDDKTYPFIKLTERHPRLIVTRNVKK 120

20 Query: 121 SDGTYFGPYPDSCGAATEIKRLDLRFPFKKCTNPANKVCFYVHLGQCNAHTVCQTNKAYW 180  
 G YFGYP+ AA E K+LLDRL+P +KC+ ++VC YYHLGQC A V ++  
 Sbjct: 121 DKGRYFGPYPNVQAARETKKLLDRLYPLRKCSKLPDRVCLYVHLGQCLAPCVKDISETN 180

25 Query: 181 DSLREDVKQFLNGKDNKIVNGLTEKMKSAAMTMEFERAAEYRDLIEAISLLRTKQORVIHQ 240  
 L E + +FL G N++ L EKM AA +EFERA E RD I I KQ++  
 Sbjct: 181 RELVESITRFLRGGYNEVKKELEKMEHAENLEFERAKELRDQIAHIESTMEKQKMTMN 240

30 Query: 241 DMKDRDVFYGFYFDKGWMCVQVFFVRNGKLIQRDVNMFFYYNEPEEDFLTYYIGQFYQDTKH 300  
 D+ DRDVF Y DKGWMCVQVFF+R GKLI+RDV+MFP Y E +E+FLT+IGQFY H  
 Sbjct: 241 DLVDRDVFAYAYDKGWMCVQVFFIRQKLIERDVSMFPLYQEADDEEFLTFIGQFYSKNNH 300

Query: 301 FLPKEVFIPQDIDAKSVETIVGCKIVKPORGKR 333  
 FLPKE+ +P ID +E ++ + +P++G +  
 35 Sbjct: 301 FLPKEILVPDSIDQSMIEQLLETNVHQPKKGPK 333

There is also homology to SEQ ID 2568.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**40 Example 2079**

A DNA sequence (GBSx2194) was identified in *S.agalactiae* <SEQ ID 6433> which encodes the amino acid sequence <SEQ ID 6434>. This protein is predicted to be maltose operon transcriptional repressor (rbsR). Analysis of this protein sequence reveals the following:

Possible site: 52

45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3761(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9393> which encodes amino acid sequence <SEQ ID 9394> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAD02112 GB:AF039082 putative maltose operon transcriptional  
 repressor [Lactococcus lactis]  
 Identities = 64/166 (38%), Positives = 105/166 (62%), Gaps = 13/166 (7%)

-2345-

Query: 1 MGKSAIDYLYKKGHKSIQFVTDDLNSEVSEERYLG YFKGARKLGLNQKPALLFDRGNPQV 60  
 +G+ A+ L + H++I FVTD +EV EERY G+ A +LGL+ LLF N +  
 Sbjct: 169 LGREAVRLLAQLNHQNI SFVTDTKETEVEERYQGFKDEAERLGLSHD--LLFMDSNFSL 226

Query: 61 LEEFINRVKEET TALIVIGD TVSVRVMQFLSFYKLVKVPDDISIMTFNNSLFSHLIHPYL 120  
 E TAL+V+ D +S++V++ L L VP+D+S++T+NNS+F +IHPYL  
 Sbjct: 227 RNE-----TALVMDDVLSLKVVERLRSQGLNVPEDVSLITYNNSIFGAMIHPYL 276

Query: 121 STFDINVNVLGR TSVRRLLIDI KSPDKVFSETIIVPFTLEERESVR 166  
 +TFDI++ LG ++++++D+ + + + +TII PF L RES +  
 Sbjct: 277 TTFDIHIEQLGASAIKKILDRLDNKENLPEKTII-PFELIVRESTK 321

There is also homology to SEQ ID 5082.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2080

A DNA sequence (GBSx2195) was identified in *S.agalactiae* <SEQ ID 6435> which encodes the amino acid sequence <SEQ ID 6436>. This protein is predicted to be 4-alpha-glucanotransferase (malQ). Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2003(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA26923 GB:J01796 amylomaltase [Streptococcus pneumoniae]  
 Identities = 250/500 (50%), Positives = 329/500 (65%), Gaps = 4/500 (0%)

Query: 1 MKKRASGVLMIHITSLPGDLGIGTFGREAYAFVDFLVETDQKFWQILPLTTTSFGDSPYQS 60  
 MKKR SGVLMHI+SLPG GIG+FG+ AY FVDFLV T Q++WQILPL TS+GDSFYQS  
 Sbjct: 1 MKKRQSGVLMHISSLP GAYGIGSFGQSAYDFVDFLVRTKQRYWQILPLGATS YGDSFYQS 60

Query: 61 FSAVAGNTHLIDFLLTLEGFISKDDYQNI SFGQDPEVVDYAGLFEKRRPVLEKAVKNFL 120  
 FSA AGNTH ID D+L +G + D + + FG D VDYA ++ RRP+LEKAVK F  
 Sbjct: 61 FSAFAGNTHFIDL DILVEQGLLEASDLEGVDFGSDASEVDYAKIYYARRP LLEKAVKRFF 120

Query: 121 QEERATRLMSDFLQE-EKWVTDFAEFMAIKEHFGNKALQEWDKAIIRREEEALAGYRQK 179  
 E + F Q+ + W+ FAE+MAIKE+F N A EW D R+ AL YR++  
 Sbjct: 121 -EVGDVDKFEKFAQDNQSWLELFAEYMAIKEYFDNLAWTEWPDADARARKASALSYREQ 179

Query: 180 LSEVIKYHEVTQYFFYKQWFLKEYANDKGIQIIGDMPYVSADSVVWTPMELFKLDRD 239  
 L++ + YH VTQYFF++QW +LK YAND I+I+GDMPYIV+ DS ++W P LFK D +  
 Sbjct: 180 LADKLVYHRVTQYFFYKQWFLKAYANDNHIEIVGDMPYIVAE DSSDMWANPHLFKTDVN 239

Query: 240 KQPLAIAIGVPADDFSDGQLWGNPIYNWDYHKESDFDWWIYRIQSGVKMYDYLRI DHFKG 299  
 + IAG P D+FS QQLWGNPIY+W+ + + WWI R++ K+YD +RIDHF+G  
 Sbjct: 240 GKATCIAGCPPDEF SVTGQLWGNPIYDWEAMDKDGYKWWIERLRESFKIYDIVRIDHFRG 299

Query: 300 FSDYWEIRGDYQTANDGSWQPAFGPELFATIKEKLGDLPIIAENLGYIDERAERLLAGTG 359  
 F YWEI TA G W PG +LFA +KE+LG+L IIAE+LG++ + L TG  
 Sbjct: 300 FESYWEIPAGSDTAAPGEWVGKPGYKLF AAVKEELGELNIIAEDLGFM TDEVIELRERTG 359

Query: 360 FPGMKIMEFGFYDTTGSIDIPHNYTENTIA YAGTHDNEVINGWFEN-LTVEQKAYAENY 418  
 FPGMKI++F F + SID PH N++ Y GTHDN + GW+ N + + Y Y  
 Sbjct: 360 FPGMKILQFAF-NPEDESIDSPHLAPANSVMYTGTHDNNTVLGWYRNEID DATREYMARY 418

-2346-

Query: 419 MRRLPNEPITETVLRITLYATVSQTTITCMQDLLDKPADSRMNPNTVGGNWQWRMRKEDL 478  
 R E + +LRT++++VS I MQDLL+ ,RMN P+T+GGNW WRM ++ L  
 Sbjct: 419 TNRKEYETVVHAMLRTVFSSVSFMAIATMQDLELDEARMNFPSTLGGNWSWRMTEDQL 478

5 Query: 479 TENRKAFLKEITTIYNRGNK 498  
 T + L ++TTIY R N+  
 Sbjct: 479 TPAVEEGLDLTTIYRRINE 498

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6437> which encodes the amino acid  
 10 sequence <SEQ ID 6438>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.85 Transmembrane 435 - 451 ( 435 - 451)

15 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1341(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 313/495 (63%), Positives = 387/495 (77%)

Query: 1 MKKRASGVLMHITSLPGDLGIGTFGREAYAFVDVFLVETDQKFWQILPLTTTSFGDSPYQS 60  
 M KRASG+LMHI+SLPG GIGTFG+ A+ FVDVFL ET Q +WQILPLTTTSFGDSPYQS  
 25 Sbjct: 1 MNKRASGILMHISLPGKFGIGTFGKSAFEFVDFLAETKQTYWQILPLTTTSFGDSPYQS 60

Query: 61 FSAVAGNTHLIDFDLLTLEGFISKDDYQNI SFGQDPEVVDYAGLFEKRRPVLEKAVKNFL 120  
 FSA+AGNTH IDF+LL + + D +I+FG +PE VDYA LF+ RRP+LEKAV+ F+  
 Sbjct: 61 FSAIAGNTHFIDFELLVDDELLEAADLCDITFGTNPEAVDYAQLFQVRRLLEKAVRAV 120

30 Query: 121 QEERATRLMSDFLQEEKWVTDFAEFMAIKEHFGNKALQEWDDKAIIRREEEALAGYRQKL 180  
 E+ L F W+TDFAEFMA+KE+F NKALQ+WDD+ +I+R+E++L YR+ L  
 Sbjct: 121 AEQENVCKLEAFETASSWLTDFAEFMALKEYFNKALQDWDDETVIKRQEDSLNNYRELL 180

35 Query: 181 SEVIKYHEVTQYFFYKQWFELKEYANDKGIQIIGDMPYVSADSVEVWVMPPELFLDRDK 240  
 ++ I YH+V QYFFY+QW LK YAN KGI+IIGDMPYVSADSVEVWVMPPELFLK+D DK  
 Sbjct: 181 AKKITYHKVCQYFFYQQWSALKTYANHKGIEIIGDMPYVSADSVEVWVMPPELFLKVDSDK 240

40 Query: 241 QPLAIAGVPADDFSDGQWLGNPIYNWDYHKSDFDWWIYRIQSGVKMYDYLRIDHFKGF 300  
 +PL IAGVPAD FS+DGQWLGNP YNW H++S+F WWIYRIQ K+YD LRIDHFKGF  
 Sbjct: 241 KPLFIAGVPADGFSEDDGQWLGNPTYNWSAHEKS NFAWWIYRIQESFKLYDQLRIDHFKGF 300

45 Query: 301 SDYWEIRGDYQTANDGSWQPAPGPELFATIKEKLGDLPIIAENLGYIDERAERLLAGTGF 360  
 SD+WEI +TA +G W APG LF+ ++E LG+LPPIAENLGYIDE+AE+LLA TGF  
 Sbjct: 301 SDFWEIPAGDKTARNGHWASAPGIALFSVREALGELPIIAENLGYIDEKAEQLLASTGF 360

50 Query: 361 PGMKIMEFGFYDTTGN SIDIPHNYTENTIAYAGTHDNEVINGWFENLTVEQKAYAENYMR 420  
 PGMKI+EFG +D T SID+PH Y N +AY GTHDNEV+NGW++NL+ EQ + NY+  
 Sbjct: 361 PGMKILEFGLFDITSQSIDLPHYDRNCVAYTGTHDNEVNGWYDNLSEEQVHFVNYYLH 420

55 Query: 421 RLPNEPITETVLRITLYATVSQTTITCMQDLLDKPADSRMNPNTVGGNWQWRMRKEDLTE 480  
 + +E IT+ +LRT++A+V T I C+QDLLDK SRMNPNT+GGNWQWRM +L +  
 Sbjct: 421 KHADESITKAMLRITFASVCDTAILCIQDLLDKGSRMNPNTTIGGNWQWRMLDGELNQ 480

Query: 481 NRKAFLKEITTIYNR 495  
 + K +L +T +Y R  
 Sbjct: 481 DHKDYLIYLTDL YGR 495

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 60 vaccines or diagnostics.

-2347-

**Example 2081**

A DNA sequence (GBSx2196) was identified in *S.agalactiae* <SEQ ID 6439> which encodes the amino acid sequence <SEQ ID 6440>. This protein is predicted to be glycogen phosphorylase (malP). Analysis of this protein sequence reveals the following:

```

5   Possible site: 40
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2678(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAC00218 GB:AF008220 glycogen phosphorylase [Bacillus subtilis]
    Identities = 297/776 (38%), Positives = 452/776 (57%), Gaps = 41/776 (5%)

    Query: 13  GKVLSELTNEEIYVELLNFFVKEEAAA-----KSKNSSQRKVYYISAEFLIGKLLSNL 65
              GK  +   + Y L N V+E +A          KS+++S ++ YY+S EFL+G+LL NL
    Sbjct: 21  GKSFKDSAKLDQYKTLGNMNVREYISADWIETNEKSRNSGKQTYLLSIEFLLGQLEQNL 80

20   Query: 66  INLGIYKDVKKELELVGKSIABIEDVEPEPSLGNGLGRLASC FIDSISLGINGEVGL 125
              +NLG+ V+ L+ +G ++ EI +E + LGNGGLGRLA+CF+DS++SL + G G+G+
    Sbjct: 81  MNLGVRDVVEAGLKEIGINLEEILQIENDAGLGNGLGRLAACFLDSLASLNLPGHGMGI 140

25   Query: 126 NYHCGLFKQVFRNNQQEAEANYWIEN-NSWLVPD-DISYDVPF-----RDFTLKSRL 175
              Y GLF+Q + Q W++N N W V D + DVPF + L R
    Sbjct: 141 RYKHGLFEQKIVDGHQVELPEQWLKGNVWEVRNADQAVDVPFWGEVHMTEKSGRLHFRH 200

30   Query: 176 DR-----IDVLGYKKDTKNYLNLFIDGLDYNLIEKGITFDKTEIKNLTFLFLYP 225
              ++ I ++GY+ T N L L++ + Y G + ++ FLYP
    Sbjct: 201 EQATIVTAVPYDIPITIGYETGTVNTLRLWNAE--PYAHYHGGNLSYKRETEAVSEFLYP 258

    Query: 226  DSDKNGELLRIYQQYFMVSNAAQLLIDEAIERGSLNHLDAEYAYVQINDTHPSMVIPEL 285
              DD+ G++LR+ QQYF+V + + +++ + +L L + + INDTHP++ +PEL
35   Sbjct: 259 DDTHEGKILRLKQQYFLVCASLSIVNNYRKTHKSLSGLHKKVSIHINDTHPALAVPEL 318

    Query: 286  IRLLTEKHGFEFDEAVSVVRNMVGYNHTTILAEALEKWPLEYLNEVVPHLVTI IKKLDQM 345
              +R+L ++ ++EA + + + YTNHT L+EALEKWP+ ++P + II+++++
40   Sbjct: 319 MRILLDEENMSWEEAWHITVHTISYTNHTLSEALEKWPIHLFKPLLPRMYMIEEINER 378

    Query: 346  IRE-----EQTNPEVQIIDEAGRVHMAHMDIHFSTSVNGVAALHTEILKNSELKVFY 397
              + E I G V MAH+ I S SVNGVA +H++ILK E++ F+
    Sbjct: 379 FCRAVWEKYPGDWKRIENMAITAGVVKMAHLAIVGSYSVNGVAKIHSIDILKEREMRDFH 438

45   Query: 398 DIYPDKFNNKTNGITFRRLWLEFANQDLADYLKELIGDSYLTDATQLEKLLTYADSNEVHD 457
              ++P++FNNKTNGI RRWL AN L+ + E IGD ++ L +L YA +
    Sbjct: 439 LLFPNRFNNKTNGIAHRRWLLKANPGLSAIITEAGDEWVKQPESLIRLEPYATDPAFIE 498

50   Query: 458 KLAAIKFKNKLALKRYLKENKGIELDEYSIIDTQIKRFHEYKQQMNALYVIHKYLEIKR 517
              + K K K L + G+ ++ SI D Q+KR H YKRQ +N L++++ Y +K
    Sbjct: 499 QFQNNKSKKKQELADLIFCTAGVVVNPEISIFDVQVKRLHAYKQQLNLVHIMLYNRLKE 558

    Query: 518  GH-FPSRKLTVIFGGKAAPAYTIAQDIHLLCLSELINNDPEVNKYNLNVHLVENYNVTV 576
              F T IFG KA+P+Y A+ II LI ++E +N DP V + + V +ENY V++
55   Sbjct: 559 DSGFSIYPQTFIFGAKASPSYYYAKKIIKLIHSVAEKVNYDPAVKQLIKVVFLENYRVSM 618

    Query: 577  AEKLIPATDISEQISLASKEASGTGNMKFMLNGALTGTMDGANVEIAELAGKENIYTFG 636
              AE++ PA+D+SEQIS ASKEASGTGNMKFM+NGALT+GT DGAN+EI E G + IYTFG
    Sbjct: 619 AERIFPASDVSEQISTASKEASGTGNMKFMMNGALTGTGHDGANIEILERVGPDCIYTFG 678

60   Query: 637 KSDSTIINLYETSGYRSKDYDKKVIKREAVDFIISDDIVSLGNAERLKRHLDELV-GKD 695
              +D +++ E GYRS++YY D+ IR+ D +I+ G A+ + + D L+ D
    Sbjct: 679 LKADEVLSYQENGGYRSREYYQHRRRIQVADQLINGFFE--GEAEFESIFDSLLPHND 736

```

-2348-

Query: 696 WFMTLIDLKEYIYAVKEQVLADYEDYESWNKKVIHNIKAGFFSSDRTIEQYNQDIW 751  
 + L D Y +E++ ADY + W++ I NIA +G+FFSSDRTI +Y +DIW  
 Sbjct: 737 EYFVLKDFSSYADAQERIQADYRERRKWEHSIVNIAHSGYFSSDRTIREYAKDIW 792

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6441> which encodes the amino acid sequence <SEQ ID 6442>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -2.71 Transmembrane 538 - 554 ( 538 - 554)

----- Final Results -----

bacterial membrane --- Certainty=0.2084(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 629/754 (83%), Positives = 696/754 (91%), Gaps = 2/754 (0%)

Query: 1 MTRNFTTYYVVGQQ-GKVLSELTNEEIYVELLNLFVKEEAAAKSKNSSQKVVYISAEFLIGK 59  
 MTR FT YV + GK L++ +NEEIY+ LLNFVKEEA+ K+KNS++RKVVYISAEFLIGK  
 Sbjct: 1 MTR-FTEYVETKLGKSLTQASNEEIYLSLLNFVKEEASHKAKNSAKRKVVYISAEFLIGK 59

Query: 60 LLSNNLINLGIYKDVKKELELVGKSIABIEDVEPEPSLNGGGLGRLASC FIDSISSSLGIN 119  
 LLSNNLINLGIYKD+K+EL GKSIAB+EDVE EPSLNGGGLGRLASC FIDSI+SLGIN  
 Sbjct: 60 LLSNNLINLGIYKDIKEELAAAGKSIABEVEDVELEPSLNGGGLGRLASC FIDSIASLGIN 119

Query: 120 GEGVGLNYHCGLFKQVFRNNQQEAEANYWIENNSWLVP TDISYDVPPRDFTLKSRLDRID 179  
 GEGVGLNYHCGLFKQVF++N+QAE N+WIE++SWLVP TDISYDVPF++FTLKSRLDRID  
 Sbjct: 120 GEGVGLNYHCGLFKQVFKHNEQAEPNFWIEDDSWLVP TDISYDVPFKNFTLKSRLDRID 179

Query: 180 VLGYKDKTKNYLNLFDIDGLDYNLIEKGITFDKTEIKKNTLFLYPDDSDKNGELLRIYQ 239  
 VLGYK+DTKNYLNLFDI+G+DY LI+ GI+FDKT+I KNLTLFLYPDDSDKNGELLRIYQ  
 Sbjct: 180 VLGYKRDTKNYLNLFDIEGVYGLIKDGISFDKTQIAKNLTLFLYPDDSDKNGELLRIYQ 239

Query: 240 QYFMVSNAQAQLLIDEAIERGSLNHLDAEYAYVQINDTHPSMVIPELIRLLTEKHGFDFDE 299  
 QYFMVSNAQAQL+IDEAIERGSLNHLDA+YAYVQINDTHPSMVIPELIRLLTEKHGF+DFDE  
 Sbjct: 240 QYFMVSNAQAQLIIDEAIERGSLNHLDAEYAYVQINDTHPSMVIPELIRLLTEKHGFDFDE 299

Query: 300 AVSVVRNMVGYTNHTILAEALEKWPLEYLNVEVPHLVTTIIKKLDQMIREEQTNPEVQIID 359  
 AV+VV+NMVGYTNHTILAEALEKW P YLNEVPHLVTTII+KLD ++R E ++P VQIID  
 Sbjct: 300 AVAVVKNMVGYTNHTILAEALEKWPTAYLNEVPHLVTTIEKLDALVRSEVSDPAVQIID 359

Query: 360 EAGRVHMAHMDIHFSTSVNGVAALHTEILKNSSELKVFDIYDPKFNNKTINGITFRRWLEF 419  
 E+GRVHMAHMDIHF+TSVNGVAALHTEILKNSSELK FYD+YP+KFNNKTINGITFRRWLEF  
 Sbjct: 360 ESGRVHMAHMDIHFATSVNGVAALHTEILKNSSELKAFYDLYPEKFNNKTINGITFRRWLEF 419

Query: 420 ANQDLADYLKELIGDSYLT DATQLEKLLTYADSNEVHDKLAAIKFKNLALKRYLKENKG 479  
 ANQDLADY+KELIGD YLT DAT+LEKL+ +AD VH KLA IKF NKLALKRYLK+NK  
 Sbjct: 420 ANQDLADYIKELIGDEYLT DATKLEKLMAFADDKAVHAKLAETKFNNLALKRYLKDNKD 479

Query: 480 IELDEYSIIDTQIKRFHEYKRQMNALYVIHKYLEIKRGHFPSRKLTVIFGGKAAPAYTI 539  
 IELDE+SIIDTQIKRFHEYKRQMNALYVIHKYLEIK+G+ P RK+TVIFGGKAAPAY I  
 Sbjct: 480 IELDEHSIIDTQIKRFHEYKRQMNALYVIHKYLEIKKGNLPRKIITVIFGGKAAPAYII 539

Query: 540 AQDIIHLILCLSELINNDPEVNKYLNVHLVENYNVTVAEKLIPATDISEQISLASKEASG 599  
 AQDIIHLILCLSELINNDPEV+ YLNVHLVENYNVTVAE LIPATDISEQISLASKEASG  
 Sbjct: 540 AQDIIHLILCLSELINNDPEVSPYLNVLHVENYNVTVAEHLIPATDISEQISLASKEASG 599

Query: 600 TGNMKFMLNGALT LGTMDGANVEIAELAGKENIYTFGKSDTTIINLYETSGYRSKDYDDK 659  
 TGNMKFMLNGALT LGTMDGANVEIAELAG ENIYTFGKSDTTIINLY T+ Y +KDYDD  
 Sbjct: 600 TGNMKFMLNGALT LGTMDGANVEIAELAGMENIYTFGKSDTTIINLYATASYVAKDYDDN 659

Query: 660 DKVIREAVDFIISDDIVSLGNAERLKRHLDELVGKDWFM TIDLKEYIYAVKEQVLADYED 719  
 I+ AV+FIIS ++++ GN ERL RL+ EL+ KDWFM TIDL+EYI VKE++LADYED  
 Sbjct: 660 HPAIKAAVNFIISPELLAFGNEERLDRLYKELISKDWFM TIDLLEEYIEVKEKMLADYED 719

-2349-

Query: 720 YESWNKKVIHNIKAGFFSSDRTEQYNQDIWHS 753  
 + W KV+HNIKAGFFSSDRTEQYN+DIWHS  
 Sbjct: 720 QDLWMTKVVHNIKAGFFSSDRTEQYNEDIWHS 753

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2082

A DNA sequence (GBSx2197) was identified in *S.agalactiae* <SEQ ID 6443> which encodes the amino acid sequence <SEQ ID 6444>. This protein is predicted to be glycerol-3-phosphatase transporter (glpT). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.88 | Transmembrane | 339 - 355 ( 333 - 359) |
| INTEGRAL | Likelihood = -8.01  | Transmembrane | 432 - 448 ( 426 - 450) |
| INTEGRAL | Likelihood = -6.74  | Transmembrane | 92 - 108 ( 91 - 127)   |
| INTEGRAL | Likelihood = -6.69  | Transmembrane | 194 - 210 ( 190 - 214) |
| INTEGRAL | Likelihood = -3.77  | Transmembrane | 367 - 383 ( 364 - 385) |
| INTEGRAL | Likelihood = -2.81  | Transmembrane | 111 - 127 ( 109 - 127) |
| INTEGRAL | Likelihood = -2.28  | Transmembrane | 407 - 423 ( 406 - 424) |
| INTEGRAL | Likelihood = -2.02  | Transmembrane | 165 - 181 ( 165 - 182) |
| INTEGRAL | Likelihood = -0.64  | Transmembrane | 29 - 45 ( 29 - 45)     |

----- Final Results -----

bacterial membrane --- Certainty=0.5352(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44575 GB:U28354 IS629 ORFB fused with sequences similar to E. coli GlpT and UhpT proteins, Swiss-Prot Accession Number P08194 and P09836; Method: conceptual translation supplied by author [Shig  
 Identities = 174/321 (54%), Positives = 241/321 (74%), Gaps = 4/321 (1%)

Query: 109 GVIPSVITSIWLFITIMYLINGWLQGMGYPPGARTLVYWDNKERIKYATIWNLSHNFPGA 168  
 GV P V + + + YL+NGW+QGMGYPPGA+TLV+WY+++ERI +AT+WNLSHN GGA  
 Sbjct: 12 GVGP-VCSELHIAPSTYLLNGWIQGMGYPPGAKTLVFWEYHRERISWATLWNLSHNVGGA 70

Query: 169 IAPILTGVGLALAGNDSLQARAAAYWPGVVACLLAVLVYFLQEDTPESIGLEPPIEEYHK 228  
 +AP+L G G+ +L+ ARAA+ FPGV+ ++VL+YF+Q D P S+GLPPIEE+  
 Sbjct: 71 LAPVLIGFSFGFGDSALDHARAAFIFFGVLCMAMSVLIYFIQVDRPVSVGLPPIEEWKG 130

Query: 229 EQYTNVVDSSDILEEPEVLGMGEIHKYILPNTKLMWASLYSIFVYILRYGIVSWTPKFL 288  
 ++ E+ L + +II+K+I+ N KL++ +Y FVYILRYGIVSW PKFL  
 Sbjct: 131 NVVSHPAKGR---EQGPRLSIPDIIRKHIIRNNKLIYCCIYGSFVYILRYGIVSWAPKFL 187

Query: 289 ATSVQDGGKGITATAGMGGFSLFEIGGIIGMLTAGYLSAKVFNKSKPLTNVAFVLVVAILL 348  
 + S+ GKG + A MGG S+FEIGG+ GML AGYLS ++F+NSKPLTN FL + I+L  
 Sbjct: 188 SDSLDVGGKDMGKLASMGGSVFEEIGGVAGMLLAGYLSVRLFRNSKPLTNTLFLALTIIIL 247

Query: 349 LAAYWFIPAGPQYMALDFIILLGLGASIYGPVMMVGLYAMELVPKAAGAASGLTGTFYSY 408  
 L AYW++P+G +Y+ L++ IL+ LG ++YGPVM +GLY+MELVPK AAGAASGL+GTFSY  
 Sbjct: 248 LIAYWYVPSGNEYLWLNITILLLGLAVYGPVMFIGLYSMELVPKEAAGAASGLSGTFSY 307

Query: 409 VGGATIATLAIGIIIDHFGWG 429  
 + G+ +ATL +G+++D+ GWG  
 Sbjct: 308 IFGSIVATLGMGLVVDYLGWG 328

-2350-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6445> which encodes the amino acid sequence <SEQ ID 6446>. Analysis of this protein sequence reveals the following:

Possible site: 36

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

|    |          |                     |               |                        |
|----|----------|---------------------|---------------|------------------------|
| 5  | INTEGRAL | Likelihood = -12.37 | Transmembrane | 185 - 201 ( 175 - 208) |
|    | INTEGRAL | Likelihood = -9.13  | Transmembrane | 114 - 130 ( 90 - 134)  |
|    | INTEGRAL | Likelihood = -7.75  | Transmembrane | 322 - 338 ( 320 - 345) |
|    | INTEGRAL | Likelihood = -6.79  | Transmembrane | 421 - 437 ( 419 - 439) |
|    | INTEGRAL | Likelihood = -6.37  | Transmembrane | 91 - 107 ( 90 - 113)   |
| 10 | INTEGRAL | Likelihood = -5.36  | Transmembrane | 163 - 179 ( 161 - 181) |
|    | INTEGRAL | Likelihood = -5.20  | Transmembrane | 350 - 366 ( 347 - 371) |
|    | INTEGRAL | Likelihood = -4.41  | Transmembrane | 23 - 39 ( 22 - 41)     |
|    | INTEGRAL | Likelihood = -3.77  | Transmembrane | 257 - 273 ( 249 - 273) |
|    | INTEGRAL | Likelihood = -1.33  | Transmembrane | 61 - 77 ( 61 - 77)     |
| 15 | INTEGRAL | Likelihood = -1.28  | Transmembrane | 383 - 399 ( 383 - 399) |
|    | INTEGRAL | Likelihood = -0.90  | Transmembrane | 299 - 315 ( 299 - 315) |

----- Final Results -----

|    |                     |                                           |
|----|---------------------|-------------------------------------------|
|    | bacterial membrane  | --- Certainty=0.5946(Affirmative) < succ> |
| 20 | bacterial outside   | --- Certainty=0.0000(Not Clear) < succ>   |
|    | bacterial cytoplasm | --- Certainty=0.0000(Not Clear) < succ>   |

The protein has homology with the following sequences in the databases:

>GP:AAF96050 GB:AE004355 glycerol-3-phosphate transporter [Vibrio cholerae]  
Identities = 128/438 (29%), Positives = 215/438 (48%), Gaps = 17/438 (3%)

|    |            |                                                                   |
|----|------------|-------------------------------------------------------------------|
| 25 | Query: 1   | LFMEEDYNKREP-EKFTQFLRRQKVFFVAFF-GYVCAYLVRNNFKLMSNTIMVQNGWDK 58    |
|    |            | LF + +R P +K R + F+ F GY YL R NF L + +++ G+ +                     |
|    | Sbjct: 21  | LFKPAHAHTQRLPSDKVDSVYSRLRWQLFIGIFVGYAGYYLGRKNFSL-AMPYLIEQGFSR 79  |
| 30 | Query: 59  | AQIAILLSCLTVSYGLAKFYMGALGDRVSLRKLFSISLGASALICILIGFF---NSSMVV 115  |
|    |            | + + L ++++YGL+KF MG + DR + R S L SAL+ GF S+                       |
|    | Sbjct: 80  | GDLGVALGAVSIAYGLSKFLMGVSDRSNPRYFLSAGLLLSALVMFCFGFMPWATGSITA 139   |
| 35 | Query: 116 | LGILLVLGCGVQALAPASQAMIANFYFPNKTRGGAIAGWNISQNMGSALLELTIALLTSM 175  |
|    |            | + ILL L G QG PA + + + K RG ++ WN++ N+G L I + +                    |
|    | Sbjct: 140 | MFILLFLNGWFQGMGWPAACGRITMVHWSRKERGEIVSVWNVAHNVGGGL----IGPIFLL 195 |
| 40 | Query: 176 | GLVVPANGNILLAFILPGVLVFLFALCCWKLGGDNPESEGLDSLRTMYGDAGESAVASEE 235  |
|    |            | GL + N + AF +P L A+ W + D P+S GL + D + S E                        |
|    | Sbjct: 196 | GLWM-FNDDWRTAFYVPAFFAVLVAVFTWLVMRDTPOQSCGLPPIIEYKNDYPDDYDKSHE 254 |
| 45 | Query: 236 | EKHNLISYQQLIWKYVFCNPSSLVAAVNVALYFVRFGIEDWMPYLSQVANMSEAHIHFA 295   |
|    |            | + ++ ++ +KYVF N L +A N +Y +R+G+ DW P+YL + + + +A                  |
|    | Sbjct: 255 | NE--MTAKEIFFKYVFNKLLWSIAIANAFVYLIRYGVLDWAPVYLKEAKHFTVDKSSWA 312   |
| 50 | Query: 296 | ISMLEWVAIPGSLVFAWLAVR-YPNMAKVGAIGLFVLAIVFVYERLTATGAPNYFLLL 354    |
|    |            | + EW IPG+L+ W++ + + + A G + + + + V VY G P +                      |
|    | Sbjct: 313 | YFLYEWAGIPGTLCCGWISDKVFKGRRAPAGILFMVLVTLAVLVY-WFNPAGNPAVDMAA 371  |
| 55 | Query: 355 | VIAGILGSLIYGPQLIVNLTINFVPLNVAGTAIGFVGVTAYLIGNMGANWLMPIADGF 414    |
|    |            | ++A +G LIYGP +++ + + P AGTA G G+ YL G + AN ++ D F                 |
|    | Sbjct: 372 | LVA--IGFLIYGPVMLIGLYALELAPKKAAGTAAGLTGLFGYLGGAANAILGYTVDFH 429    |
|    | Query: 415 | GWFWSYIVVAALSAFSAV 432                                            |
|    |            | GW ++V+ A S +                                                     |
|    | Sbjct: 430 | GWGGGFVVLVASCVL 447                                               |

An alignment of the GAS and GBS proteins is shown below.

|    |                                                                           |                                                                |
|----|---------------------------------------------------------------------------|----------------------------------------------------------------|
| 60 | Identities = 117/439 (26%), Positives = 203/439 (45%), Gaps = 27/439 (6%) |                                                                |
|    | Query: 23                                                                 | KYPRYRVQVLISIFVGYMGYYFVRNTTSILSGILNMS---ATEIGIITCASYIAYGLSK 78 |
|    |                                                                           | ++ R + V F GY+ Y VRN ++S + + +I I+ ++YGL+K                     |
|    | Sbjct: 17                                                                 | QFLRRQKVFFVAFFGYVCAYLVRNNFKLMSNTIMVQNGWDKAQIAILLSCLTVSYGLAK 76 |



-2351-

Query: 79 FISGLISDESNSKIFLPVGLFLTGLVNVLIGVIPSVITSIWLFTIMYLINGWLQGMGYPP 138  
 F G + D + + + L + L+ +LIG S S+ + I+ ++ G +QG P  
 Sbjct: 77 FYMGALGDRVSLRKLFSSISLGASALICILIGFFNS---SMVVLGILLVLCGVVQGALAPA 133

5 Query: 139 GARTLVYWDNKRERIKYATIWNLSHNFGGAIAP-----LTGVGLALAGNDSINQARAAYW 194  
 + ++ NK R WN+S N G A+ P+ LT +GL + N ++ A+  
 Sbjct: 134 SQAMIANYPNKRTRGGAIAGWNISQNMGSALLPLTIALLTSMGLVVPANGNI---LLAFL 190

10 Query: 195 FPGVVACLLAVLVYFLQEDTPESIGLPPPIEYHKEQYTNVVDSSDILEEPEVLGMGEIIG 254  
 PGV+ L A+ + L D PES GL + + + + V S EE L ++I  
 Sbjct: 191 IPGVLVFLFALCCWKLGGDNPESEGLDSLRTMYGDAGESAVASE---EEKHNLSYQLIW 247

15 Query: 255 KYILPNTKLMWASLYSIFVYILRYGIVSWTPKFLATSVQDGGKGITATAGMGGFSLFEIG 314  
 KY+ N L+ + ++ +Y +R+GI W P +L+ I S+ E  
 Sbjct: 248 KYVFCNPSSLLLVAAVNVALYFVRFGIEDWMPYLSQVANMSEAHIHFA-----ISMLEWV 302

20 Query: 315 GIIGMLTAGYLSAKVFNKSKPLTNVAFLLVVAILLAAAYWFIPAG-PQYMALDFIILLG-L 372  
 I G L +L+ + + + V+A ++ G P Y L +++ G L  
 Sbjct: 303 AIPGSLVFVAWLAVRYPNKMMAVGAIGLFVLAIVFYERLTATGAPNYFLL--LVIAGIL 360

25 Query: 373 GASIYGPVMVGLYAMELVPKAAAGAASGLTGTFSYVGGATIATLAIGIIDDHFGWGVAF 432  
 G+ IYGP ++V + + VP AG A G G +Y+ G A + I+ D FGW ++  
 Sbjct: 361 GSLIYGPQLIVNILTINFVPLNVAGTAIGFVGVTAYLIGNMGANWLMPIADGFGWFWSY 420

Query: 433 IIF-GISGFAAIVCTLLSR 450  
 I+ +S F+A+ +L++  
 Sbjct: 421 IVVAALSAFSAVGYLILAK 439

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 30 vaccines or diagnostics.

### Example 2083

A DNA sequence (GBSx2198) was identified in *S.agalactiae* <SEQ ID 6447> which encodes the amino acid sequence <SEQ ID 6448>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3202(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6449> which encodes the amino acid sequence <SEQ ID 6450>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 45 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4473(Affirmative) < succ>  
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 54/100 (54%), Positives = 67/100 (67%)

55 Query: 1 MTYELCLEYGTYPPLRPVDAWADEINTAPAFITEDKKLLELLEEVNTLFHELFLTIESCFH 60  
 MTYELCLEYGTYP L VDA+ E P FI ED+ L LE +N LFH+LF+TIE FH  
 Sbjct: 1 MTYELCLEYGTYP LSRVDAYWGEDQNPPFTIQEDRLCHKLETMNHLFHDLFVTIESQFH 60

-2352-

Query: 61 YIGHDFPEKRAKITQIYHVIEHLSIHYPEYDIKIESLLM 100  
 Y+G + PEKRA+I +Y + L Y +Y IKIE+ L+  
 Sbjct: 61 YVGFNMPEKRAQIRILYQEVATILSKYKDYPKIETFL 100

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2084

A DNA sequence (GBSx2199) was identified in *S.agalactiae* <SEQ ID 6451> which encodes the amino acid sequence <SEQ ID 6452>. Analysis of this protein sequence reveals the following:

10 Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 15 bacterial cytoplasm --- Certainty=0.2369(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAB81912 GB:U92974 unknown [Lactococcus lactis]  
 Identities = 213/322 (66%), Positives = 260/322 (80%), Gaps = 5/322 (1%)

Query: 1 MSEKIRVLLYYKYVSIENAEYYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60  
 M++ RVLLYY+YV IE+ E +A KHL CK +GLKGRIL+ADEGINGTVSG E T Y  
 Sbjct: 1 MTQDYRVLLYYQYVPIEDGETFAQKHLADCKELGLKGRILVADEGINGTVSGTIEQTINAY 60

25 Query: 61 MDVWHSDEFADLWFKIDEENQQAFRKMFVRYKKEIVHLGLEDDNNFDSNDINPLETTGEYL 120  
 M+ + +D RF+ FKIDE Q AF+KM VRY+ E+V+L LED D+NPLE TG YL  
 Sbjct: 61 MELMKNDPRFSSTIFKIDEAEQNAFKMHVRYRPELVNLSLED-----DVNPLELTGAYL 115

30 Query: 121 NPKQFKEALLDEDTVLDTRNDYEYDLGHFRGAIRPDIRNFRELPOWVRDNKDKFMEKRV 180  
 +PK+F+EA+LDE+TVV+D RNDYE+DLGHFRGAIRP+IR+FRELPQW+RDNK++FMEKRV  
 Sbjct: 116 DPKEFREAMLDETVVIDARNDYEFDLGHFRGAIRPEIRSFRELPOWIRDNKEQFMEKRV 175

35 Query: 181 VVYCTGGVRCCKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGLWDGAMYVFDDRISVP 240  
 + YCTGG+RCEKFSGW+VREGFKDVGQL GGIATYGKDPEVQG+LWDG MYVFD RI+VP  
 Sbjct: 176 LTYCTGGIRCEKFSGWLVRGFKDVGQLLGGIATYGKDPEVQGLWDGQMYVFDRI+VP 235

Query: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENEAKYVRGCSPECRAHERNRY 300  
 IN ++ +D+FDG+PCERY+NC NP CN+Q+ ASEENEAKY+ CS ECR H NRY  
 40 Sbjct: 236 INQKEHVIVGRDWFDPSPCERYINCNPENRQMLASEENEAKYLGACSHECRVHPNNRY 295

Query: 301 VQENGLSRQEWAEERLEAIGESL 322  
 ++ + LS QE ERL + + L  
 45 Sbjct: 296 IKAHQSLNQEVQERLALLEKDL 317

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6453> which encodes the amino acid sequence <SEQ ID 6454>. Analysis of this protein sequence reveals the following:

50 Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 55 bacterial cytoplasm --- Certainty=0.2443(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 321/324 (99%), Positives = 323/324 (99%)

Query: 1 MSEKIRVLLYYKYVSIENAEYYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60

-2353-

Sbjct: 1 MSEKIRVLLYYKYVSIENA+EYAAKHLEFCKSIGLKGRILIADEGINGTVSGDYETTQKY 60  
 Query: 61 MDVWHSDEFADLWFKIDEENQQA FRKM FVRYKKEIVHLGLEDDNNFSDINPLETTGEYL 120  
 MDVWHSDEFADLWFKIDEENQQA FRKM FVRYKKEIVHLGLEDDNNFSDINPLETTGEYL  
 Sbjct: 61 MDVWHSDEFADLWFKIDEENQQA FRKM FVRYKKEIVHLGLEDDNNFSDINPLETTGEYL 120  
 Query: 121 NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFREL PQWVRDNKDKFMEKRV 180  
 NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFREL PQWVRDNKDKFMEKRV  
 Sbjct: 121 NPKQFKEALLDEDTVVLDTRNDYEYDLGHFRGAIRPDIRNFREL PQWVRDNKDKFMEKRV 180  
 Query: 181 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGE LWDGAMYVFDDRISVP 240  
 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGE LWDGAMYVFDDRISVP  
 Sbjct: 181 VVYCTGGVRCEKFSGWMVREGFKDVGQLHGGIATYGKDPEVQGE LWDGAMYVFDDRISVP 240  
 Query: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENEAKYVRGCSPECRAHERNRY 300  
 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENE KYVRGCSPECRAHERNRY  
 Sbjct: 241 INHVNPTVISKDYFDGTPCERYVNCANPFCNKQIFASEENETKYVRGCSPECRAHERNRY 300  
 Query: 301 VQENGLSRQEWAEERLEAIGESLPQ 324  
 VQENGLSRQEWAEERLEAIGESLP+  
 Sbjct: 301 VQENGLSRQEWAEERLEAIGESLPE 324

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2085

A DNA sequence (GBSx2200) was identified in *S.agalactiae* <SEQ ID 6455> which encodes the amino acid sequence <SEQ ID 6456>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have an uncleavable N-term signal seq  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC83954 GB:L47648 putative [Bacillus subtilis]  
 Identities = 54/192 (28%), Positives = 89/192 (46%), Gaps = 14/192 (7%)  
 Query: 5 QTIIIGAGAAGIGFGSAMQRLGLTNFLIEKGHIGESFLRWPRTTQFITPSFTTNGFGFP 64  
 + IIIG G G+ +++G+ + L+IEKG++ S +P F + S  
 Sbjct: 5 KAIIGGGPCGLSAAIHLKQIGI-DALVIEKGNVNSIYNYPHTQTFFSSSEKLE----- 58  
 Query: 65 DLNAVIPDTSPAFSFEKEHLSGVEYARYLQLVAAHYNLPIQNETSVLSIDK-RDSL FVIK 123  
 I D AF E ++ Y + V N+ + V + K +++ FVI+  
 Sbjct: 59 -----IGDV--AFITENRKPVRIQALSYYREVVKRKNIRVNAFEMVRKVTKTQNTTFVIE 111  
 Query: 124 TSKGDFSADYLIMATGEFQNPNTIDIKGADLGMHYGQVDNFIKSDNPFIIIGGNESACD 183  
 TSK ++ Y I+ATG + +PN + + G DL + H D ++IGG S+ D  
 Sbjct: 112 TSKETYTPYCIATGYDHPNMGVPGEDLPKVFHYFKEGHPYFDKDVVIGGKNSSVD 171  
 Query: 184 ALTHLVYLGNOV 195  
 A LV G +V  
 Sbjct: 172 AALELVKSGARV 183

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2354-

A related GBS gene <SEQ ID 8973> and protein <SEQ ID 8974> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1   Crend: 2
McG: Discrim Score:      5.05
GvH: Signal Score (-7.5): -3.14
    Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 0 value:  0.26 threshold:  0.0
    PERIPHERAL Likelihood =  0.26      6
    modified ALOM score:  -0.55

*** Reasoning Step: 3

----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

33.2/56.1% over 281aa

Bacillus subtilis
EGAD|109228| hypothetical protein Insert characterized
GP|2635109|emb|CAB14605.1||Z99117 alternate gene name: yrdP Insert characterized
GP|1934657|gb|AAB80908.1||U93876 hypothetical protein YrdP Insert characterized
PIR|E69725|E69725 potassium uptake trkA - Insert characterized

ORF01799(310 - 1128 of 1725)
EGAD|109228| S2656(2 - 283 of 345) hypothetical protein { acillus subtilis}
GP|2635109|emb|CA 14605.1||Z99117 alternate gene name: yrdP { acillus subtilis}
GP|1934657|gb|AA 80908.1||U93876 hypothetical protein YrdP { acillus subtilis}
PIR|E69725|E69725 potassium uptake trkA - acillus subtilis
%Match = 6.1
%Identity = 33.2 %Similarity = 56.0
Matches = 77 Mismatches = 88 Conservative Sub.s = 53

270      300      330      360      390      417      444      474
YYC*LVKYFILHIYFCQGEDMKHYQTIIIGAGAAGIGFGSAMQRLGLTNFLIEKGH-IGESFL-RWPRTTQFITPSFTT
      |||:||||| ||| |:: |:::| |::| |:: |:::
40      MYDTIVIGAGQAGISIGYYLKQ-SDQKFIILDKSHEVGESWKDRYDSLVLFTSRMYSS
              10      20      30      40      50

480      510      540      570      600      630      660      690
-----NGFGFPLNAVIPDTSFAFSFEKEHLSGVEYARYQLVAAHYNLPIQNETSVLSIDKRDSLFVIKTSKGFDS
      ||| |:: ||: ||: |::| |::| |::| |::| |::|
45      LPGMHLEGEKHGFPKNEIV-----AYLKKYVKKFEIPIQLRTEVISVLKIKNYFLIKTNREEYQ
              70      80      90      100      110

720      750      822      852      882      912
ADYLI MATGEFQNPNTIDIKGADLG-----MHYGQVDNF-HIKSDNPFIIIGGNESACDALTHLVYLCNQVELYTDTFGR
      |::||| |::| |::| |::| |::| |::| |::| |::|
50      TKNLVIATGPFHTNIPNIS-KDLSDNINQLHSSQYKNSKQLAYGNVLVVGGSNGA-----
              130      140      150      160      170

942      969      996      1026
KESNPDPSPISLS-PLTKERLKHQ-DHKKEYYSISEGKKAI--EIKQIG-----
      :: |::|:: :: :| |:: |::| |::| |::|
55      -----QIAVELSKERVITYLACSNKL VYFPLMIGKRSIFWFDKLGVLHASHTSIVGKFIQKKGDPVFGHELKHAIK
              180      190      200      210      220      230      240

1068      1098      1128      1158      1188      1218      1248
-----KQYQVTFDDGSTAESFHKPILSTGFLNTCHLIDGIALFEYDKNQLPVITDEDESTIVNNCFLIGPSL
      ||::| |::| |::| |::| |::| |::| |::|
60      QKEIILKKRVIAAKQNEIIFKDSSTLE-VNNIIWATGFRNPLCWINIKGVLDQEGRIIHRGVSPVEGLYFICLPWQHQR
              260      270      280      290      300      310      320

```

-2355-

SEQ ID 8974 (GBS284) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 10; MW 42.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 9; MW 67.6kDa).

GBS284-GST was purified as shown in Figure 225, lane 7.

## 5 Example 2086

A DNA sequence (GBSx2201) was identified in *S.agalactiae* <SEQ ID 6457> which encodes the amino acid sequence <SEQ ID 6458>. This protein is predicted to be NrgA-like protein. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have an uncleavable N-term signal seq

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -11.73 | Transmembrane | 7 - 23 ( 1 - 31)       |
| INTEGRAL | Likelihood = -6.42  | Transmembrane | 86 - 102 ( 82 - 108)   |
| INTEGRAL | Likelihood = -6.42  | Transmembrane | 324 - 340 ( 318 - 342) |
| INTEGRAL | Likelihood = -5.26  | Transmembrane | 210 - 226 ( 207 - 229) |
| INTEGRAL | Likelihood = -5.10  | Transmembrane | 113 - 129 ( 112 - 133) |
| INTEGRAL | Likelihood = -1.49  | Transmembrane | 246 - 262 ( 246 - 263) |
| INTEGRAL | Likelihood = -1.17  | Transmembrane | 183 - 199 ( 183 - 199) |
| INTEGRAL | Likelihood = -0.43  | Transmembrane | 41 - 57 ( 41 - 57)     |
| INTEGRAL | Likelihood = -0.00  | Transmembrane | 265 - 281 ( 265 - 282) |

----- Final Results -----

bacterial membrane --- Certainty=0.5692(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9997> which encodes amino acid sequence <SEQ ID 9998> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15668 GB:Z99122 ammonium transporter [Bacillus subtilis]

Identities = 105/378 (27%), Positives = 181/378 (47%), Gaps = 41/378 (10%)

Query: 3 VKKGLFVFLLLCILSMWLMIFGVAFFYFGLH-QSLTSRIIYQFVLTVLLTTTAWFMGAY 61  
++ G VF+ C L +WLM G+A +Y G + +++ S ++ F ++ + + W + Y  
Sbjct: 1 MQMGDTVFMMFFCALLVWLMTPGLALFYGGMVKSKNVLSTAMHSFS-SIAIVSIVVWVLFY 59

Query: 62 FLAFEGHFKTIVFQFQEQADGKQI-----VNCLFQLCFALYAVVMLIGSIIDR 107  
LAF + + A K + + +FQ+ FA+ ++ G+ +R  
Sbjct: 60 TLAFAPGNSIIGGLEWAGLKGVGFDPGDYSDTIPHSLFMMFQMTFAVLTTAISGAFER 119

Query: 108 VQTKRLLAVVSWLFLVYTPPLAYLIWNSEGVFAGMGVLDGSGMIVHLSAGLSSYILAHV 167  
++ LL V W LVYTP+A+ +W G ++G LDF+GG +VH+S+G++ +LA V  
Sbjct: 120 MRFGAFLLFSVLWASLVYTPVAHWVWGG-GWIGQLGALDFAGGNVHISGAVGLVLAIV 178

Query: 168 IGK-----SEHQHNKVKNDSLFLGMILITFGWFGFNMGPVGEWNSQAIMILNTIFAIG 222  
+GK + HN + FLG LI FGWFGFN+G + A+ +NT A  
Sbjct: 179 LGKRKDGTAASSPHNLIYT---FLGGALIWFVGWFGFNVGSALTLDGVAMYAFINTINTAAAA 235

Query: 223 GGLAWTLAAKWNGEETGSLNLGIIIVGLVTSTAGVGYLLTQQLLAVTFFASLFTYFVTD 282  
G W L ++ ++G I GLV T G++ + + + ++  
Sbjct: 236 GIAGWILVEWIINKKPTMLGAVSGAIGLVAITPAAGFVTPFASIIIGIIGGAVCFWGVF 295

Query: 283 YVAKAFAIDDDVSSFGMNGIGGLGSLGVLGFLKLSHMP-----VQLLAL 326  
+ K F DD + +FG++GIGG G + GLF + + Q++A+  
Sbjct: 296 SLKKKFGYDDALDAFGLHGIGGTWGGIATGLFATTSVNSAGADGLFYGDASLIWKQIVAI 355

Query: 327 ATTILLSIIMTYIISKAI 344  
A T + I+T++I K +  
Sbjct: 356 AATYVVFVIVTFVVIKIV 373

-2356-

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8975> and protein <SEQ ID 8976> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1   Crend: 4
      McG: Discrim Score:      17.19
      GvH: Signal Score (-7.5): -4.07
      Possible site: 24
      >>> Seems to have an uncleavable N-term signal seq
10     ALOM program count: 9 value: -11.73 threshold: 0.0
      INTEGRAL Likelihood = -11.73 Transmembrane 7 - 23 ( 1 - 31)
      INTEGRAL Likelihood = -6.42 Transmembrane 86 - 102 ( 82 - 108)
      INTEGRAL Likelihood = -6.42 Transmembrane 324 - 340 ( 318 - 342)
      INTEGRAL Likelihood = -5.26 Transmembrane 210 - 226 ( 207 - 229)
15     INTEGRAL Likelihood = -5.10 Transmembrane 113 - 129 ( 112 - 133)
      INTEGRAL Likelihood = -1.49 Transmembrane 246 - 262 ( 246 - 263)
      INTEGRAL Likelihood = -1.17 Transmembrane 183 - 199 ( 183 - 199)
      INTEGRAL Likelihood = -0.43 Transmembrane 41 - 57 ( 41 - 57)
      INTEGRAL Likelihood = -0.00 Transmembrane 265 - 281 ( 265 - 282)
20     PERIPHERAL Likelihood = 0.26 152
      modified ALOM score: 2.85

      *** Reasoning Step: 3

25     ----- Final Results -----
      bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30     The protein has homology with the following sequences in the databases:

      ORF01800(307 - 1332 of 1641)
      EGAD|19589|BS3646(1 - 373 of 404) probable ammonium transporter {Bacillus subtilis}
      OMNI|NT01BS4254 ammonium transporter SP|Q07429|NRGA_BACSU PROBABLE AMMONIUM TRANSPORTER
      (MEMBRANE PROTEIN NRGA). GP|143264|gb|AAA17399.1||L03216 membrane-associated protein
35     {Bacillus subtilis} GP|1684645|emb|CAB05374.1||Z82987 unknown {Bacillus subtilis}
      GP|2636176|emb|CAB15668.1||Z99122 ammonium transporter {Bacillus subtilis}
      PIR|A36865|A36865 ammonium transporter nrgA - Bacillus subtilis
      %Match = 13.5
      %Identity = 30.0 %Similarity = 54.8
40     Matches = 104 Mismatches = 149 Conservative Sub.s = 86

      144      174      204      234      264      294      324      354
      PFSMIRKVFVSPNRCMAEKPPIPAAPAPIIMV**CFMSSP*QK*MCKIKYLT*S*Q*YSLTNKRVFVKGLFVFLLLCILSM
      :: | | |::: | | :
45     MQMGDTVMFFCALLV
      10

      384      411      441      471      501      531
      WLMIFGVAFYFYGSLH-QSLTSRIIYQFVLTVLLTTTAWFMGAYFLAFEGHFKTVFQFQBADGKQI-----
50     : ||| |::: | | : : : | : : : : | | | : : : | | :
      WLMTPLGALFYGGMVKSKNVLTAMHSF-SSIAIVSIVVWLFGYTLAFAPGNSIIGGLEWAGLKGVGFDPGDYSDTIPHS
      30      40      50      60      70      80      90

      579      609      639      669      699      729      759      789
      VNCLFQLCFALYAVVMLIGSIIDRVQTKRLLAVVSWLFLVYTPPLAYLIWNSEGVFAKMGVLDVDFSGGMIVHLSAGLSSYI
55     : ||| |::: | : : : : || | | |::: | : : | | |::: | : : : :
      LFMMFQMTFAVLTTAISGAFARMRFGAFLLFVWASLVYTPVAHWVWGG-GWIGQLGALDFAGGNVVHISSGVAGLV
      110      120      130      140      150      160      170

      819      849      873      903      933      963      993      1023
      LAHVIGKSEHQHNKVKNDLSF--LGMILITFGWFGFNMGPVGEWNSQAIMILLNTIFAIIGGLAWTLAAKWNGBEEKTC
60     || |::: | : : : || | | |::: | : : : | | | : : : | : : :
      LAIVLGKRKDGTAASSPHNLIYTFGLGALIWFVGFGFNVGSALTLDGVAMYAFINTNTAAAGIAGWIL-VEWIIKKPTM
      190      200      210      220      230      240      250

```

-2357-

```

1050      1080      1110      1140      1170      1200      1230      1260
-SLLNGIIVGLVLTSTAGVGYLLTWQLLAVTFFASLFTYFVTDYXAKAFAIDDVSSFGMNGIGLLGSLGVGLFKLSHMP
5  : : | | | | | : : : : : : : | | : : | : | : | : : :
LGAVSGAIAGLVAITPAAGFVTPFASIIIGIIGGAVCFWGVFSLKKKFGYDDALDAFGLHGIGGTWGGIATGLFATTSVN
      270      280      290      300      310      320      330

1272      1302      1332      1362      1392      1422      1452
V-----QLLALATTILLSIIMTYIISKAI FRK**IRLRCTSQPYLLF*QGE*LNRIINH FHY*TLSXX*
10  : : | : | : : | : : | : : | : :
SAGADGLFYGDASLIWKQIVAI AATYVFV FIVTFV IIKIVSLFLPLRATEEEEESLGLDLTMHGEKAYQDSM
      350      360      370      380      390      400

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2087

A DNA sequence (GBSx2202) was identified in *S. agalactiae* <SEQ ID 6459> which encodes the amino acid sequence <SEQ ID 6460>. This protein is predicted to be dUTPase (dut). Analysis of this protein sequence reveals the following:

```

20  Possible site: 51
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.2731(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9471> which encodes amino acid sequence <SEQ ID 9472> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

30  >GP:CAA72644 GB:Y11901 dUTPase [Lactococcus lactis]
    Identities = 67/144 (46%), Positives = 90/144 (61%), Gaps = 8/144 (5%)

35  Query: 40  RGFELVSQFSNKELLPKRETAHAAGYDLKVAKKTVIEPGEITLVPTGIKAYMQPGEVLYL 99
        RGF+   +      +P+R T H+AGYD+  ++   I+P EI +V TG+   +   EVL L
    Sbjct: 3   RGFK---KLDGNATIPERATKHSAGYDISASEVTTIQPD EIKMVSTGLAVQLGDDEV LKL 59

    Query: 100 YDRSSNPRKKGIVLINSVGVIDGDYNNQVNEGHIFAQMQNITDQAVILEEGERIVQAVF 159
        YDRSSNP K+GI LINSVG+ID DYY   +      NI+ + V + +G+RI+Q VF
40  Sbjct: 60 YDRSSNPVKRGIALINSVGIIIDSDYYPQEFK-----GLFMNISKEPVTISKQQRIMQGVF 114

    Query: 160 APFLLADDDQATGMRTGGFGSTGK 183
        +L   DDD A G RTGGFGSTG+
45  Sbjct: 115 VKYLTIDDDNANGKRTGGFGSTGE 138

```

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6461> which encodes the amino acid sequence <SEQ ID 6462>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 37
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
55      bacterial cytoplasm --- Certainty=0.2519(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-2358-

Identities = 115/148 (77%), Positives = 125/148 (83%)

Query: 36 MSKVRGFELVSQFSNKEKLPKRETAHAAGYDLKVAKKTVEPGEITLVPTGIKAYMQPGE 95  
 M+K+RGFELVS F+N +LLPKRET HAAGYDL VA+ I PGEI LVP TG+KAYMQ GE  
 5 Sbjet: 1 MTKIRGFELVSSFTNPDLKPKRETHAAGYDLVAEAVTIAPGEIKLVPTGVKAYMQDGE 60

Query: 96 VLYLYDRSSNPRKKGIVLINSVGVIDGDYNNQVNEGHIFAQMQNITDQAVILEEGERIV 155  
 VLYLYDRSSNPRKKGI+LINSVGVID DYY N+ NEGHIFAQMQNITD V L GERIV  
 10 Sbjet: 61 VLYLYDRSSNPRKKGIILINSVGVIDADYYGNEANEGHIFAQMQNITDHPVTLAVGERIV 120

Query: 156 QAVFAPFLLADDDQATGMRTGGFGSTGK 183  
 Q VF PFL+AD DQA G RTGGFGSTG+  
 Sbjet: 121 QGVFMPFLIADGDQARGERTGGFGSTGQ 148

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2088

A DNA sequence (GBSx2203) was identified in *S.agalactiae* <SEQ ID 6463> which encodes the amino acid sequence <SEQ ID 6464>. This protein is predicted to be RadA homolog (radA). Analysis of this  
 20 protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2628(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB11863 GB:Z99104 DNA repair protein homolog [Bacillus subtilis]  
 Identities = 285/453 (62%), Positives = 358/453 (78%), Gaps = 4/453 (0%)

Query: 1 MAKKKSVFTCQECGYQSPKYLGRCPNCSAWSSFVEEVEVQEVKNARVSLNGEKS RPTKLK 60  
 MAK KS F CQ CGY+SPK++G+CP C AW++ VEE+ + N R + + K  
 35 Sbjet: 1 MAKTKSKFICQSCGYESPKNMGKCPGCGAWNTMVEEMIKKAPANRRAAFSSHVQTVQKPS 60

Query: 61 DVSSINYS---RTKTDMEFNRLVGGGVVPGSLVLIGGDPGIGKSTLLLQVSTQLA-NKG 116  
 ++SI S R KT + EFNRLVGGGVV GSLVLIGGDPGIGKSTLLLQVS QL+ +  
 40 Sbjet: 61 PITSIETSEEP RVK TQLGEFNRLVGGGVVKGSLVLIGGDPGIGKSTLLLQVSAQLSGSSN 120

Query: 117 TVLYVSGEESAEQIKLRSERLGDIDNEFYLYAETNMQSIRSEIEKIKPDFLIIDSITQTIM 176  
 +VLY+SGEES +Q KLR++RLG + ++ +ET+M+ I S I+++ P F+++DSIQT+  
 Sbjet: 121 SVLYISGEESVKQTKLRADRLGINNPSLHVLSETDMEYISSAIQEMNPSFVVVDSITQTVY 180

45 Query: 177 SPEVSSVQGSVSQVREVTAEMLQAKTNNIATFIVGHVTKETLAGPRMLEHMDTVLYF 236  
 +++S GSVSQVRE TAEML++AKT I FIVGHVTKEG++AGPR+LEHMDTVLYF  
 Sbjet: 181 QSDITSAPGSVSQVRECTAELMKIAKTGIPFIVGHVTKEGSIAGPRLEHMDTVLYF 240

Query: 237 EGERHHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLERLDGATGSAIVVT 296  
 EGERHHTFRILRAVKNRFGSTNE+GIFEM+ GL EVLNPS++FLEER G+ GS+I +  
 50 Sbjet: 241 EGERHHTFRILRAVKNRFGSTNEMGIFEMREEGLTEVLNPSSEIFLEERSAGSAGSSITAS 300

Query: 297 MEGTRPILAEVQALVPTVFGNAKRTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSA 356  
 MEGTRPIL E+QAL++PT FGN +R TG+D NRVSL+MAVLEKR GLLLQNQDAYLK A  
 55 Sbjet: 301 MEGTRPILVEIQALISFTSFGNPRMATGIDHNRVSLIMAVLEKRVGLLLQNQDAYLKVA 360

Query: 357 GGVKLDEPAIDLAVAVAIASSYKEKPTNPQESFGEIGLTGEIRRVTRIEQRINEASKLG 416  
 GGVKLDEPAIDLA+ ++IASS+++ P NP + FIGE+GLTGE+RRV+RIEQR+ EA+KLG  
 60 Sbjet: 361 GGVKLDEPAIDLAIVISIASFRDTPNPADCFIGEVLGTGEVRRVSRIEQRVKEAAKLG 420

Query: 417 FTKIYAPKNSLAGIEIPKGIDVIGVTTVSQVLK 449



-2359-

F ++ P +L G PKGI+VIGV V++ L+  
 Sbjct: 421 FKRMIIPAANLDGWTKPKGIEVIGVANVAEALR 453

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6465> which encodes the amino acid sequence <SEQ ID 6466>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.2191(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 416/453 (91%), Positives = 441/453 (96%)

Query: 1 MAKKKSVFTCQECGYQSPKYLGRCPNCSAWSSFVEEVEVQEVKNARVSLNGEKS RPTKLK 60  
 MAKKK+ F CQECGYQSPKYLGRCPNCSAWSSFVEEVEV+EVKNARVSL GEKSRP KLK  
 Sbjct: 1 MAKKKATFICQECGYQSPKYLGRCPNCSAWSSFVEEVEVKEVKNARVSLAGEKSRPVKLK 60

20 Query: 61 DVSSINYSRTKTDMDDEFNRVLGGGVVPGSLVLIGGDPGIGKSTLLLQVSTQLANKGTVLY 120  
 DV +I+Y RT+TDM EFNRLVGGGVVPGSL+LIGGDPGIGKSTLLLQVSTQLANKGTVLY  
 Sbjct: 61 DVDNISYHRTQTDMSEFNRLVGGGVVPGSLILIGGDPGIGKSTLLLQVSTQLANKGTVLY 120

25 Query: 121 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQSIRSEIEKIKPDFLIIDSIQTIMSPEV 180  
 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQ+IR+EIE IKPDFLIIDSIQTIMSP++  
 Sbjct: 121 VSGEESAEQIKLRSERLGDIDNEFYLYAETNMQAIRTEIENIKPDFLIIDSIQTIMSPDI 180

30 Query: 181 SSVQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKETLAGPRMLEHMVDTVLYFEGER 240  
 + VQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKETLAGPRMLEHMVDTVLYFEGER  
 Sbjct: 181 TGVQGSVSQVREVTAELMQLAKTNNIATFIVGHVTKETLAGPRMLEHMVDTVLYFEGER 240

35 Query: 241 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSAIVVTMEGT 300  
 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGSA+VVTMEG+  
 Sbjct: 241 HHTFRILRAVKNRFGSTNEIGIFEMQSGGLVEVLNPSQVFLEERLDGATGS AVVTMEGS 300

40 Query: 301 RPILAEVQALVTPTVFGNAKRTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK 360  
 RPILAEVQ+LVTPTVFGNA+RTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK  
 Sbjct: 301 RPILAEVQSLVTPTVFGNARRTTTGLDFNRVSLIMAVLEKRCGLLLQNQDAYLKSAGGVK 360

45 Query: 361 LDEPAIDLAVAVAIASSYKEKPTNPQESFIGEIGLTGEIRRVTRIEQRINEASKLGFTKI 420  
 LDEPAIDLAVAVAIASSYKEKPT+PQE+F+GEIGLTGEIRRVTRIEQRINEA+KLGFTK+  
 Sbjct: 361 LDEPAIDLAVAVAIASSYKEKPTSPQEAFLGEIGLTGEIRRVTRIEQRINEAAKLGFTKV 420

45 Query: 421 YAPKNSLAGIEIPKGIDVIGVTTTVSQVLKAVFS 453  
 YAPKN+L GI+IP+GI+V+GVTTV QVL AVFS  
 Sbjct: 421 YAPKNALQGIDIPQGIIEVGVTTTVGQVLNAVFS 453

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2089

A DNA sequence (GBSx2204) was identified in *S.agalactiae* <SEQ ID 6467> which encodes the amino acid sequence <SEQ ID 6468>. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 60 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2360-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA97750 GB:Z73419 hypothetical protein Rv1284 [Mycobacterium
tuberculosis]
Identities = 69/162 (42%), Positives = 100/162 (61%), Gaps = 2/162 (1%)

Query: 3 TYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRNAG 62
T D++L N YA LP+ P +AIV CMD+RL V + LG+ G+AH++RNAG
Sbjct: 2 TVTDDYLANNVDYASGF-KGPLEMPPSKHIAIVACMDARLDVYRMLGIKEGEAHVIRNAG 60

Query: 63 GRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQFTTNEAFAAQLQRDLGVDMHGHDFLP 122
VTDDV+RSL ISQ+ LGTREI++LHHTDCG TFT++ F +Q + G+
Sbjct: 61 CVVTDDVIRSLAISQRLLGTRERILLHHTDCGMLTFTDDDFKRAIQDETGIRPTWSP-ES 119

Query: 123 FNDIEESVREDVAKLHASPLIPDDVVISGAIYDVTGRMVEV 164
+ D E VR+ + ++ +P + + G ++DV TG++ EV
Sbjct: 120 YPDAVEDVRQSLRRIEVNPFFVTKHTSLRGFVFDVATGKLINEV 161
```

There is also homology to SEQ ID 6470:

```
Identities = 126/164 (76%), Positives = 146/164 (88%)

Query: 1 MTTYFDNFLKTNQAYADLHGTAHLPIKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60
+ +YF++F+ NQAY LHGTAHLP+KPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN
Sbjct: 1 LMSYFEHFMAANQAYVALHGTAHLPLKPKTKVAIVTCMDSRLHVAQALGLALGDAHILRN 60

Query: 61 AGGRVTDDVLRSLVISQQQLGTREIVVLHHTDCGAQFTTNEAFAAQLQRDLGVDMHGHDF 120
AGGRVT+D++RSLVISQQQ+GTREIVVLHHTDCGAQFTTNE FA + LGVD+ G DF
Sbjct: 61 AGGRVTEDMIRSLVISQQQMGTRERIVVLHHTDCGAQFTTNEGFAKHIHEHLGVDVSGQDF 120

Query: 121 LPFNDIEESVREDVAKLHASPLIPDDVVISGAIYDVTGRMVEV 164
LPF D+E+SVRED+AK+ AS LI DDVVI+GA+YDVTG+M +V
Sbjct: 121 LPFQDVEDSVREDMAKIRASSLISDDVINGAVYDVTGKMTQV 164
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2090

A DNA sequence (GBSx2205) was identified in *S.agalactiae* <SEQ ID 6471> which encodes the amino acid sequence <SEQ ID 6472>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0536(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9473> which encodes amino acid sequence <SEQ ID 9474> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC73407 GB:AE000137 putative oxidoreductase [Escherichia coli K12]
Identities = 199/438 (45%), Positives = 286/438 (64%)

Query: 1 MKKYDVIVLGFSGKAGKTLAAKLATQGKSVAMVEEDDKMYGGTCINIGCIPTKTLVLSASK 60
M KY +++GFGKAGKTLA LA G VA++E+ + MYGGTCINIGCIPTKTL+ A +
Sbjct: 10 MNKYQAVIIGFGKAGKTLAVTLAKAGWRVALIEQSNAMYGGTCINIGCIPTKTLVHDAQQ 69

Query: 61 NHDFQEAMITRNEVTSRLRAKNFAMLDNKTVDVYNKARFISNKVVELTGGADKQELTA 120
+ DF A+ +NEV + LR KNF L + +DV + +A FI+N + + E+
```

-2361-

Sbjct: 70 HTDFVRAIQRKNEVVNFLRNKNFHNLADMPNIDVIDGQAEFINNHSLRVHRPEGNLEIHG 129

Query: 121 DVIIINTGAKSVQLPIPLGLADSQHVVDSTAIQELAHLPKRLGIIGGGNIGLEFATLYSEL 180  
+ I INTGA++V PIPG+ + VVDST + L LP LGI+GGG IG+EFA++++

5 Sbjct: 130 EKIFINTGAQTVPPIPGITTTFGVVDSTGLLNLKELPGHLGILGGGYIGVEFASMFANF 189

Query: 181 GSKVTVIDSQSRIFAREEEELSEMAQDYLEEMGISFKLSADIKSVQNEDEDVVISFEDEK 240  
GSKVT++++ S RE+ ++++ L + G+ L+A ++ + + + V + E +

10 Sbjct: 190 GSKVTILEAASLFLPREDRDIADNIATILRDQGVDIILNAHVERISHHENQVQVHSEHAQ 249

Query: 241 LSFDAVLYATGRKPNTTEGLALENTDIKLTERRGAIADVEYCQTSVENIFAVGDVNGGPQFT 300  
L+ DA+L A+GR+P T L EN I + ERGAI VD+ T+ +NI+A+GDV GG QFT

Sbjct: 250 LAVDALLIASGRQPATASLHPENAGIAVNERGAIVVDKRLHTTADNIWAMGDVTGGGLQFT 309

15 Query: 301 YISLDDSRIVLNYLNCDDYSLKNRGAVPTSTFTNPPLATVGLDEKTAKEKGQVKSNSL 360  
YISLDD RIV + L + S +R VP S F PPL+ VG+ E+ A+E G ++ +L

Sbjct: 310 YISLDDYRIVRDELLGEGKRSTDDRNVPYSVFMTPPLSRVGMTEEQARESGADIQVVTL 369

Query: 361 LVSAIPRAHVNDLRGIFKVVDTETNLILGARLFGAESHELINIITMAMDNKIPYTYFQ 420  
V+A+PRA V ND RG+ K +VD +T +LGA L +SHE+INI+ M MD +PY+ +

20 Sbjct: 370 PVAaiprarmndtrgvLKAIVDNKTQRM LGASLLCVDSEMINIVKMVMDAGLPYSILR 429

Query: 421 KQIFTHPTMVENFNDLFN 438  
QIFTHP+M E+ NDLF+

25 Sbjct: 430 DQIFTHPSMSLSLNDLFS 447

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 30 Example 2091

A DNA sequence (GBSx2206) was identified in *S.agalactiae* <SEQ ID 6473> which encodes the amino acid sequence <SEQ ID 6474>. This protein is predicted to be glutamyl-tRNA synthetase (gltX). Analysis of this protein sequence reveals the following:

Possible site: 43

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

          bacterial cytoplasm --- Certainty=0.2245(Affirmative) < succ>

          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

40           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9475> which encodes amino acid sequence <SEQ ID 9476> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10953> which encodes amino acid sequence <SEQ ID 10954> was also identified.

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC31971 GB:U49789 glutamyl-tRNA synthetase [Bacillus subtilis]  
Identities = 273/491 (55%), Positives = 353/491 (71%), Gaps = 19/491 (3%)

Query: 20 LANKIRVRYAPSPTGLLHIGNARTALFNLYLARHHGGDFVIRIEDTDRKRHVEDGERSQL 79  
+ N++RVRYAPSPTG LHIGNARTALFNLY+AR+ GG F+IR+EDTD+KR++E GE+SQL

50 Sbjct: 1 MGNEVRVRYAPSPTGHLHIGNARTALFNLYLFARNQGGKFIIIRVEDTDKKRNIEGGEQSQL 60

Query: 80 ENLRWLGMWDWDESPET---HENYRQSERLELYQRYIDQLLAEGKAYKSYVTEELAAERE 136  
L+WLGDWDES + + YRQSER ++Y+ Y ++LL +G AYK Y TEEEL ERE

55 Sbjct: 61 NYLKWLGIDWDESVDVGGEYGPYRQSERNDIYKVYYEELLEKGLAYKCYCTEEELEKERE 120

Query: 137 RQELAGETPRYINEFIGMSETEKEAYIAEREAGIIPTVRLAVNESGIYKWTDMVKGDI 196  
Q GE PRY + +++ E+E +IAE G P++R V E + + D+VKG+I

-2362-

Sbjct: 121 EQIARGEMPRYSKGRDLTQEEQEKFAE----GRKPSIRFRVPEGKVIAFNNDIVKGEIS 176

Query: 197 FE SGNIGGDWVIQKKDGYPTYNFAVVIDDHDMQISHVIRGDDHIANTPKQLMVYEALGWE 256  
 FE IG D+VI KKD G PTYNFAV IDD+ M+++HV+RG+DHI+NTPKQ+M+Y+A GW+

5 Sbjct: 177 FESDGIG-DFVIVKKDGTPTYNFAVAIDDLMMKMTHTVLRGEDHISNTPKQIMITYQAFGWD 235

Query: 257 APQFGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYMSEAVFNFIALLGWNPGGEEEI 316  
 PQFGHMTLI+N E+ KKL SKRD + +QFIE Y++ GY+ EA+FNFI LLGW+P GEEE+

10 Sbjct: 236 IPQFGHMTLIVN-ESRKKLSKRDESIIQFIEQYKELGYLPEALFNFIGLLGWSPVGEEEL 294

Query: 317 FSREQILNLFDENRLSKSPAADFQKKMDWMSNDYLNADNFESVFALCKPFLEEAGRL--- 373  
 F++EQ I +FD NRLSKSPA FD K+ W++N Y+K D + V L P L++AG++

Sbjct: 295 FTKEQFIEIFDVNRLSKSPALFDMHKLKWNQYVKKLDLDQVVELTLPHLQKAGKVGTE 354

15 Query: 374 -----TDKAEKLVELYQPOLKSADEIVPLTDLFFADFPPELTEAEKEVMAAETVPTVLSAF 428  
 + KL+ LY QL EIV LTDLFF D E + K V+ E VP VLS F

Sbjct: 355 LSAAEQEWKRKLISLYHEQLSYGAIEVELTDLFFTDIEYNQEAQAVLEEEQVPEVLSTF 414

Query: 429 KEKLVSLSDDEEFTTRDTIFPQIKAVQKETGIKGNLFMPIRIAVSGEMHGPPELPTIYLLG 488  
 KL L EEFT D I IKAVQKETG KGK LFMPIR+AV+G+ HGPPEL +I L+G

20 Sbjct: 415 AAKLEEL--EEFTPDNIKASIKAVQKETGHKGGKLFMPIRVAVTGQTHGPPELQSIELIG 472

Query: 489 KEKSVQHIDNM 499  
 KE ++Q + N+

25 Sbjct: 473 KETAIQRLKNI 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6475> which encodes the amino acid sequence <SEQ ID 6476>. Analysis of this protein sequence reveals the following:

Possible site: 24

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1966 (Affirmative) < succ>

35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 434/481 (90%), Positives = 459/481 (95%)

40 Query: 20 LANKIRVRYAPSPTGLLHIGNARTALFNLYARHHGGDFVIRIEDTDRKRHVEDGERSQL 79  
 ++ IRVRYAPSPTGLLHIGNARTALFNLYAR HGG F+IRIEDTDRKRHVEDGERSQL

Sbjct: 1 MSKPIRVRYAPSPTGLLHIGNARTALFNLYARRHGGTFIIRIEDTDRKRHVEDGERSQL 60

Query: 80 ENLRWLGMWDDESPETHENYRQSERLELYQRYIDQLLAEGKAYKSYVTEELAAERERQE 139  
 ENL+WLGMWDDESPETHENYRQSERL LYQ+YIDQLLAEGKAYKSYVTEELAAERERQE

45 Sbjct: 61 ENLKLWLGMWDDESPETHENYRQSERLALYQQYIDQLLAEGKAYKSYVTEELAAERERQE 120

Query: 140 LAGETPRYINEFIGMSETEKEAYIAEREAAGIIPVRLAVNESGIYKWTD MVKGDIEFEG 199  
 AGETPRYINEFIGMS EK YIAEREAAGI+PTVRLAVNESGIYKWTD MVKGDIEFEG

50 Sbjct: 121 AAGETPRYINEFIGMSADEKAKYIAEREAAGIVPTVRLAVNESGIYKWTD MVKGDIEFEG 180

Query: 200 SNIGGDWVIQKKDGYPTYNFAVVIDDHDMQISHVIRGDDHIANTPKQLMVYEALGWEAPQ 259  
 NIGGDWVIQKKDGYPTYNFAVV+DDHDMQISHVIRGDDHIANTPKQLMVYEALGWEAP+

55 Sbjct: 181 GNIGGDWVIQKKDGYPTYNFAVVDDHDMQISHVIRGDDHIANTPKQLMVYEALGWEAPE 240

Query: 260 FGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYMSEAVFNFIALLGWNPGGEEEIFSR 319  
 FGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYM EAVFNFIALLGWNPGGEEEIFSR

Sbjct: 241 FGHMTLIINSETGKKLSKRDNTLQFIEDYRKKGYMPEAVFNFIALLGWNPGGEEEIFSR 300

60 Query: 320 EQLINLFDENRLSKSPAADFQKKMDWMSNDYLNADNFESVFALCKPFLEEAGRLTDKAEK 379  
 EQLI LFDENRLSKSPAADFQKKMDWMSN+YLK+ADFE+V+ALCKPFLEEAGRLT+KAEK

Sbjct: 301 EQLIALFDENRLSKSPAADFQKKMDWMSNEYLKHADFETVYALCKPFLEEAGRLTEKAEK 360

Query: 380 LVELYQPOLKSADEIVPLTDLFFADFPPELTEAEKEVMAAETVPTVLSAFKEKLVSLSDDE 439  
 LVELY+POLKSADEI+PLTDLFF+DFPELTEAEKEVMA ETV TVL AFK KL ++SDE+

65

-2363-

Sbjct: 361 LVELYKPKLSADEIIPLTDLFFSDFPELTEAEKEVMAGETVSTVLQAFKAKLEAMSDDED 420

Query: 440 FTRDTIFPQIKAVQKETGIKGNLFMPIRIAVSGEMHGPELPDTIYLLGKEKSVQHIDNML 500  
 F + IFPQIKAVQKETGIKGNLFMPIRIAVSGEMHGPELP+TIYLLG++KS++HI NML

5 Sbjct: 421 FKPENIFFQIKAVQKETGIKGNLFMPIRIAVSGEMHGPELPNTIYLLGRDKSIEHIKNML 481

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2092

10 A DNA sequence (GBSx2207) was identified in *S.agalactiae* <SEQ ID 6477> which encodes the amino acid sequence <SEQ ID 6478>. This protein is predicted to be d-ribose-binding protein precursor, fragment (rbsB). Analysis of this protein sequence reveals the following:

Possible site: 24  
 >>> May be a lipoprotein

15 ----- Final Results -----

|                     |     |                              |         |
|---------------------|-----|------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.0000 (Not Clear) | < succ> |
| bacterial outside   | --- | Certainty=0.0000 (Not Clear) | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear) | < succ> |

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15613 GB:Z99122 ribose ABC transporter (ribose-binding protein) [Bacillus subtilis]  
 Identities = 143/301 (47%), Positives = 205/301 (67%), Gaps = 1/301 (0%)

25 Query: 14 MSIVLILGACGKTGLGNSSGNSTKNVTKKSADLKLGVSTTNNPYFVAMKDGIDKYAS 73  
 +S++L L T K + K+ +G+S+ST NNP+FV++K GI+K A  
 Sbjct: 5 VSVILTLSLFLLTACSLLEPPQWAKPSNSGNKKEFTIGLSVSTLNNPFFVSLKKGIEKEAK 64

30 Query: 74 NKKISIKVADAQDDAARQADDVQNFISQNVDAILINPVDSKAIVTAIKSANNANIPVILM 133  
 + + + + DAQ+D+++Q DV++ I Q VDA+LINP DS AI TA++SAN +PV+ +  
 Sbjct: 65 KRGMKVIVDAQNDSKQTSQVDEDLIQQGVDAALLINPTDSSAISTAVESANAVGVPVVTI 124

35 Query: 134 DRGSEGGKVLTTVASDNVAAGKMAADYAVKKLGGKAKAFELSGVPGASATVDRGKGFHVS 193  
 DR +E GKV T VASDNV G+MAA + KLGK AK EL GVPASAT +RG GFH++  
 Sbjct: 125 DRSABEQGVETLVASDNVKGGEAAAFAIDKLGKGAKEVLEGVPGASATRERGSFGFHN 184

40 Query: 194 AKSKLDILSSQSANFDRAKALNTTQNMIQGHKDVQIIIFAQNDEMAGAAQAVKSAGLQNV 253  
 A KL +++ QSA+FDR K L +N++QGH D+Q +FA NDEMALGA +A+ S+G +++  
 Sbjct: 185 ADQKLQVVTQKQSAFDRTKGLTVMENLLQGHDPDIQAVFAHNDEMAGALEAINSSG-KDI 243

45 Query: 254 LIVIGIDGQPDHAIDAIKKGDISATIAQQPAKMGELIAIQAAIDYKGGKVEKETISPIYLVTK 314  
 L++G DG DA +IK +SAT+AQQP +G++A +AA D GKKV+K +P+ L T+  
 Sbjct: 244 LVIGFDGNKDALASIKDRKLSATVAQQPELIGKLATEAADLHGKKVQKTISAPLKLETQ 304

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6478 (GBS203) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 12; MW 36.8kDa).

GBS203-His was purified as shown in Figure 208, lane 8.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2364-

**Example 2093**

A DNA sequence (GBSx2208) was identified in *S.agalactiae* <SEQ ID 6479> which encodes the amino acid sequence <SEQ ID 6480>. This protein is predicted to be galactoside ABC transporter, permease protein (rbsC). Analysis of this protein sequence reveals the following:

```

5   Possible site: 14
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -11.15   Transmembrane   63 - 79 ( 52 - 85)
      INTEGRAL    Likelihood = -3.66    Transmembrane  111 - 127 ( 110 - 128)
      INTEGRAL    Likelihood = -2.71    Transmembrane  168 - 184 ( 168 - 188)
10   INTEGRAL    Likelihood = -2.44    Transmembrane  189 - 205 ( 188 - 205)
      INTEGRAL    Likelihood = -0.80    Transmembrane   17 - 33 ( 17 - 33)

    ----- Final Results -----
      bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
15   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9287> which encodes amino acid sequence <SEQ ID 9288> was also identified.

20 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:CAB15612 GB:Z99122 ribose ABC transporter (permease) [Bacillus subtilis]
    Identities = 144/211 (68%), Positives = 182/211 (86%), Gaps = 1/211 (0%)

    Query: 1  MGMLNGLFISYGKLPFFIVTLATMTIFRGATLVYSNGNPITAGLSDSFLFQFLGQGYIVG 60
25   +GM+NGL I+ GK+APFI TLATMT+FRG TLVY++GNPIT GL ++ FQ G+GY +G
    Sbjct: 113 LGMINGLLITKGKMAPFIATLATMTVFRGLTLVYTDGNPIT-GLGTNYGFQMFGRGYFLG 171

    Query: 61  IPFPVILMFLTFIILYILLHKTAFGKSVYALGGNEKAAYISGIKLNKVKIIITYTISGIMA 120
30   IP P I M L F+IL++LLHKT FG+ YA+GGNEKAA ISGIK+ +VK++IY+++G+++
    Sbjct: 172 IPVPAITMVLAFVILWVLLHKTTPFGRRTYAIGNEKAALISGIKVTRVKVMIYSLAGLLS 231

    Query: 121 SISGLIITSRLSSAQPTAGASYEMDAIAAVVLGGTSLSGGKGRIIGTLIGALIIGVLNNG 180
35   +++G I+TSRL SAQPTAG SYE+DAIAAVVLGGTSLSGG+GRI+GTLIG LIIG LNNG
    Sbjct: 232 ALAGAILTSRLHSAQPTAGESYELDAIAAVVLGGTSLSGGRIVGTLLIGVLIIGTLNNG 291

    Query: 181 LNIIGVSAFWQQVVKGIVILMAVLLDRFKVA 211
      LN++GVS+F+Q VVKGIVIL+AVLLDR K A
    Sbjct: 292 LNLLGVSSFYQLVVKGIVILIAVLLDRKKSA 322

```

40 A related GBS gene <SEQ ID 8977> and protein <SEQ ID 8978> were also identified.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2094**

45 A DNA sequence (GBSx2209) was identified in *S.agalactiae* <SEQ ID 6481> which encodes the amino acid sequence <SEQ ID 6482>. Analysis of this protein sequence reveals the following:

```

    Possible site: 35
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.12    Transmembrane   75 - 91 ( 74 - 91)
      INTEGRAL    Likelihood = -0.64    Transmembrane  96 - 112 ( 96 - 112)
50

    ----- Final Results -----
      bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
55   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 5 Example 2095

A DNA sequence (GBSx2210) was identified in *S.agalactiae* <SEQ ID 6483> which encodes the amino acid sequence <SEQ ID 6484>. This protein is predicted to be ribose transport ATP-binding protein rbsa (rbsA). Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.00      Transmembrane  401 - 417 ( 401 - 417)

----- Final Results -----
      bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15611 GB:Z99122 ribose ABC transporter (ATP-binding protein)
20      [Bacillus subtilis]
      Identities = 297/493 (60%), Positives = 375/493 (75%), Gaps = 1/493 (0%)

Query: 1  MKIDMRNISKSFGTNKVLEKIDLELQSGQIHALMGENGAGKSTLMNILTGLFPASTGTIY 60
      M+I+M++I K+FG N+VL + +L G++HALMGENGAGKSTLMNILTGL A G I
25      Sbjct: 1  MQIEMKDIHKTFGKNQVLSGVSFQLMPGEVHALMGENGAGKSTLMNILTGLHKADKGQIS 60

Query: 61  IDGEERTFSNPQEAEFEFGISFIHQEMNTWPMTVLENLFLGREIKTTFGLLNQKLMRQKA 120
      I+G E FSNP+EAE+ GI+FIHQE+N WPEMTVLENLF+G+EI + G+L + M+ A
30      Sbjct: 61  INGNETYFSNPKEAEQHGIAFIHQELNIWPEMTVLENLFIGKEISSKLGVLQTRKMKALA 120

Query: 121 LETFKRLGVTIPLDIPIGNLSVGGQQQMIEIAKSLNQLSILVMDEPTAALTDTRETNLFR 180
      E F +L V++ LD G SVGQQQMIEIAK+L+ +++MDEPTAALT+RE LF
35      Sbjct: 121 KEQFDKLSVSLSDQEAGECSVGQQQMIEIAKALMTNAEVIIMDEPTAALTEREISKLFE 180

Query: 181 VIRGLKQEGVGVVYISHRMEEIFKITDFVTVMRDGVIVDTKETSLTNSDELVKKMVGRKL 240
      VI LK+ GV +VYISHRMEEIF I D +T+MRDG VDT S T+ DE+VKKMVGR+L
40      Sbjct: 181 VITALKKNQVSIVYISHRMEEIFAICDRITIMRDGKTVDTTNISETDFDEVVKKMVGREL 240

Query: 241 EDYYPEKHSEIGPVAFEVSNL-CGDNFEDVSFYVRKGEILGFSGLMGAGRTEVMRTIFGI 299
      + YP++ +G FEV N +FEDVSFYVR GEI+G SGLMGAGRTE+MR +FG+
45      Sbjct: 241 TERYPKRTPSLGDKVFEVKNASVKGSFEDVSFYVRSGEIVGVSGLMGAGRTEMMRALFGV 300

Query: 300 DKKKSGKVKIDDQEITITTPSQAIKQGIGFLTENRKDEGLILD FN IKDNMTLPSTKDFSK 359
      D+ +G++ I ++ I P +A+K+G+GF+TENRKDEGL+LD +I++N+ LP+ FS
50      Sbjct: 301 DRLDTGEIWIAGKKTAKNPQEAVKKGLGFITENRKDEGLLLDTSIRENIALPNLSSSFSP 360

Query: 360 HGFFDEKTSTTFVQQLINRLYIKSGRPDLEVGNLSGGNQKQVVLAKWIGIAPKVLILDEP 419
      G D K FV LI RL IK+ P+ +LSGGNQKQVV+AKWIGI PKVLILDEP
55      Sbjct: 361 KGLIDHKREAEFVDLLIKRLTIKTASPETHARHLSGGNQKQVVIKAWIGIGPKVLILDEP 420

Query: 420 TRGVDVGAKREIYQLMNELADRGVPIVMVSSDLPEILGVSDRIMVMHEGRISGELS RKEA 479
      TRGVDVGAKREIY LMNEL +RGV I+MVSS+LPEILG+SDRI+V+HEGRISGE+ +EA
60      Sbjct: 421 TRGVDVGAKREIYTMNELTERGVAIIMVSSSELPEILGMSDRIIVVHEGRISGEIHAREA 480

Query: 480 DQEKVMQLATGGK 492
      QE++M LATGG+
65      Sbjct: 481 TQERIMTLATGGR 493

```

There is also homology to SEQ ID 4678.

-2366-

SEQ ID 6484 (GBS407d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 2-4; MW 72kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 5 & 6; MW 47kDa).

GBS407d-His was purified as shown in Figure 235, lane 9-10.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2096

A DNA sequence (GBSx2211) was identified in *S.agalactiae* <SEQ ID 6485> which encodes the amino acid sequence <SEQ ID 6486>. This protein is predicted to be high affinity ribose transport protein rbsd (rbsD). Analysis of this protein sequence reveals the following:

Possible site: 14  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2673(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB15610 GB:Z99122 ribose ABC transporter (membrane protein)  
[Bacillus subtilis]  
Identities = 74/131 (56%), Positives = 95/131 (72%), Gaps = 1/131 (0%)  
  
Query: 1 MKKTGILNSHLAKLADDLGHTDRVCIGDLGLPVPNGIPKIDLSLTSGIPSFQEVLDIYLE 60  
25 MKK GILNSHLAK+ DLGHTD++ I D GLPVP+G+ KIDL SL G+P+FQ+ + E  
Sbjct: 1 MKKHGILNSHLAKILADLGHTDKIVIADAGLPVDPGV LKIDLSLKPGLPAFQDTAAVLAE 60  
  
Query: 61 NILVEKVILAEIEKEANPDQLSRLLAKLDNSVSI EYVSHNHLKQMTQDVKAVIRTGENTP 120  
+ VEKVI A EIK +N + ++ L L + IEY+SH K +T+D KAVIRTGE TP  
30 Sbjct: 61 EMAVEKVIAAAEIKASNQEN-AKFLENLFSEQEIEYLSHEEFKLLTKDAKAVIRTGEFTP 119  
  
Query: 121 YSNIILQSGVI 131  
Y+N ILQ+GV+  
35 Sbjct: 120 YANCILQAGVL 130

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2097

40 A DNA sequence (GBSx2212) was identified in *S.agalactiae* <SEQ ID 6487> which encodes the amino acid sequence <SEQ ID 6488>. This protein is predicted to be ribokinase (rbsK). Analysis of this protein sequence reveals the following:

Possible site: 47  
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50



-2367-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15609 GB:Z99122 ribokinase [Bacillus subtilis]
Identities = 132/293 (45%), Positives = 177/293 (60%), Gaps = 4/293 (1%)

5   Query: 1   MSNIVIIGSISMDLVMETNRIAKEGETVFGQRFMSVPGGKGANQAVAIGRLSQERDNITI 60
      M NI +IGS SMDLV+ +++ K GETV G F VPGGKGANQAVA RL + + +
      Sbjct: 1   MRNICVIGSCSMDLVVTSDKRPKAGETVLGTSFQTVPGGKGANQAVAAARLGAQ---VFM 57

10  Query: 61  LGAIGEDSFGPILLDNLNKNHVTTFDVGITIP-SSSGVAQITLYNNDNRRIIYCPGANGKVD 119
      +G +G+D +G +L+NL N V TD++ + + SG A I L DN I+ GAN +
      Sbjct: 58  VGKVGDDHYGTAILNNLKANGVRTDYMEPVTHTESGTAHIVLAEGDNSIVVVGANDDIT 117

      Query: 120 TKKWSQEWSIIKEADLVVLQNEIPHQANMKIANFCKEHSIKLLYNPAPSRETDIEMLDKV 179
      I++ D+V++Q EIP + ++ +C H I ++ NPAP+R E +D
15  Sbjct: 118 PAYALNALEQIEKQVDMVLIQQEIPETVDEVCKYCHSHDIPILNPAPARPLKQETIDHA 177

      Query: 180 DYFTPNHEHCQELFPNQKLEDILATYPEKLIIVTLGTKGAIYSKGESHLIPALETKAVDT 239
      Y TPNHEH LFP + + LA YP KL +T G +G YS G + LIP+ + VDT
20  Sbjct: 178 TYLTPNEHEASILFPELTISEALALYPAKLFITEGKQGVRYSGSKEVLIPSFPPEPVD 237

      Query: 240 TGAGDTFNAGFYAISKKFKIAKALRFATLAAHLSVQKFGAQGGMPTIKEMED 292
      TGAGDTFN AF A+++ I ALRFA AA LSV FGAQGGMPT E+E+
      Sbjct: 238 TGAGDTFNAAFAVALAEGKDIEAALRFANRAASLSVCSFGAQGGMPTRNEVEE 290
```

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2098

30 A DNA sequence (GBSx2213) was identified in *S.agalactiae* <SEQ ID 6489> which encodes the amino acid sequence <SEQ ID 6490>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35  bacterial cytoplasm --- Certainty=0.2272(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

40 A related GBS nucleic acid sequence <SEQ ID 9477> which encodes amino acid sequence <SEQ ID 9478> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15608 GB:Z99122 transcriptional regulator (LacI family)
[Bacillus subtilis]
Identities = 141/327 (43%), Positives = 204/327 (62%), Gaps = 4/327 (1%)

45  Query: 13  MSTIRQVAEKAGVSTSTVSRYSISQNGYVSQKASQKIEQAIRELHYVPNFLAQSLTKKNQ 72
      M+TI+ VA AGVS +TVSR ++ NGYV ++ ++ A+ +L+Y PN +A+SL ++++
      Sbjct: 1   MATIKDVAGAAGVSVATVSRNLNDNGYVHEETRTRVIAAMAKLNYYPNNEVARSLYKRESR 60

50  Query: 73  LVGLLLPDISNPFFPRLARGVEEFLKEQGYRVMLGNTNNKSHLEEEYLNVLQSNAAAGII 132
      L+GLLLPDI+NPFFP+LARG E+ L +GYR++ GN++ + E EYL Q++ AGII
      Sbjct: 61  LIGLLLPDITNPPFPQLARGAEDELNREGYRLIFGNSDEELKKELEYLQTFKQNHVAGII 120

      Query: 133 --TTHDFTKNHPEIDIPVVVVDRVNQETQYGVFSDNKEGGKLAAQAIWTAGATNILLIRG 190
      T + + + ++ PVV +DR E V SD G KLAAQAI + I L+RG
55  Sbjct: 121 AATNYPDLLEYSGMNYPVVFLDR-TLEGAPSVSSDGYTGKLAQAIIHGKSQRITLLRG 179

      Query: 191 PLDKADNLNQRFGSQNYLLNKGACFAIEDSASFDAEIQIEAKTLLDHHPDIDSIAPS 250
```

-2368-

P RF G+ L F + ++ASF + Q AK L +P D +IA +  
 Sbjct: 180 PA-HLPTAQDRFNGALEILKQAEVDFQVIETASFSIKDAQSMAKELFASYPATDGVIASN 238

Query: 251 DIHAIAYLHEILNRGKRIPEDVQIIGYDDILMSQFIYPSLSTIHQSSYIMGQKAAELIFK 310  
 DI A A LHE L RGK +PED+QIIGYDDI S ++P LSTI Q +Y MG++AA+L+  
 Sbjct: 239 DIQAAAVLHEALRRGKNVPEDIQIIGYDDIPQSGLLFPPLSTIKQPAYDMGKEAAKLLLG 298

Query: 311 ITNQLPITNKRIKLPVHYVERETLRRK 337  
 I + P+ I++PV Y+ R+T R++  
 Sbjct: 299 IIKQPLAETAIQMPVTYIGRKTTRKE 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6491> which encodes the amino acid sequence <SEQ ID 6492>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1657(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 232/328 (70%), Positives = 274/328 (82%)

Query: 10 GVSMSTIRQVAEKAGVSTSTVSRYISQNGYVSQKASQKIEQAIRELHYVPNFIAQSLKTK 69  
 G +M TI+QVAE+AGVS STVSRYISQ GYVS A KI+ AI +LHY PN LAQSLKTK  
 Sbjct: 14 GKAMVTIKQVAEEAGVSRSTVSRYISQKGYVSDDARHKIKAAIAKLHYTPNVLAQSLKTK 73

Query: 70 KNQLVGLLLPDISNPFFPRLARGVEEFLKEQGYRVM LGNTNNKSHLEEEYLVNLLQSNAA 129  
 KNQLVGLLLPDISNPFFPRLARG EE+LKE+GYRVM LGN ++ LEEY++VLLQSNAA  
 Sbjct: 74 KNQLVGLLLPDISNPFFPRLARGAEEY LKEGYRVM LGNISDSEALEEEYVHVLLQSNAA 133

Query: 130 GIITTHDFTKNHPEIDIPVVVVDVRVNETQYGVFSDNKEGGKLAQAIAWTAGATNILLIR 189  
 GIITTHDFTK +P + IPVVVVDVRV+QETQYGVFSDN+ GG LAAQ +W AGA +LLIR  
 Sbjct: 134 GIITTHDFTKRYPTLAIPVVVVDVRVDQETQYGVFSDNRAGGLLAAQTWVQAGAKEVLLIR 193

Query: 190 GPLDKADNLNQRFGSQNYLLNKGACFAIEDSASFDAEIQIEAKTLLDHHPDIDSIAP 249  
 GPLD A+N+N+RF+ S +YL + + DS +FDF IQ+EA L +P IDSIIAP  
 Sbjct: 194 GPLDNAENINERFEASFSYLQKQDVTMYVCDNQNFDFESIQLSEASYNLKCYPITIDSIAP 253

Query: 250 SDIHAIAYLHEILNRGKRIPEDVQIIGYDDILMSQFIYPSLSTIHQSSYIMGQKAAELIF 309  
 SDIHAIA Y+HE+ ++GK+IP+DVQIIGYDDILMSQFIYPSLSTIHQSSY+MG+ AAEL++  
 Sbjct: 254 SDIHAIA YIHELHSQKKIPQDVQIIGYDDILMSQFIYPSLSTIHQSSYLMGRYAAELVY 313

Query: 310 KITNQLPITNKRIKLPVHYVERETLRRK 337  
 I +QL + RIKLPVHYVERET+R++  
 Sbjct: 314 TIASQLTVKANRIKLPVHYVERETIRKR 341

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2099

A DNA sequence (GBSx2214) was identified in *S.agalactiae* <SEQ ID 6493> which encodes the amino acid sequence <SEQ ID 6494>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -13.80 | Transmembrane | 27 - 43 ( 24 - 51)     |
| INTEGRAL | Likelihood = -10.61 | Transmembrane | 337 - 353 ( 329 - 362) |
| INTEGRAL | Likelihood = -9.18  | Transmembrane | 257 - 273 ( 249 - 276) |
| INTEGRAL | Likelihood = -8.92  | Transmembrane | 302 - 318 ( 291 - 326) |

-2369-

----- Final Results -----

bacterial membrane --- Certainty=0.6519(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8979> which encodes amino acid sequence <SEQ ID 8980> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 6

SRCFLG: 0

McG: Length of UR: 4

Peak Value of UR: 3.20

Net Charge of CR: 1

McG: Discrim Score: 6.06

GvH: Signal Score (-7.5): 0.0500002

Possible site: 46

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

Amino Acid Composition: calculated from 47

ALOM program count: 3 value: -10.61 threshold: 0.0

INTEGRAL Likelihood = -10.61 Transmembrane 326 - 342 ( 318 - 348)

INTEGRAL Likelihood = -9.18 Transmembrane 246 - 262 ( 238 - 265)

INTEGRAL Likelihood = -8.92 Transmembrane 291 - 307 ( 280 - 315)

PERIPHERAL Likelihood = 4.98 152

modified ALOM score: 2.62

icml HYPID: 7 CFP: 0.525

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5246(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]

Identities = 103/352 (29%), Positives = 191/352 (54%), Gaps = 9/352 (2%)

Query: 15 AWKELTFYKKKYLIELLIIVMMFMVFLSGLANGLGRAVSAAIENNPQTYILNEGAEQ 74

A +EL K + LLI ++ ++ FMV L+GL GL R ++ + + PAQ+++ + A+

Sbjct: 4 ALRELQHQKLRLSLIGGIVALIAFMVFMLTGLTRGLSRDSASLLDTPAQSFVTTKEADG 63

Query: 75 VITSSVLTKDQTDNLNLKDSSTTLNIQRSSLTRQGHEKKIDISYFAIDKDSFMAPTSL 134

V+ S L+ + +++L + ++ ++ +K++ +D F+AP +S

Sbjct: 64 VLNRSFSLPEQ---VSALQQDNEDAAFAQTFFVSFSGDKQLSGVLLGVDPGRGFLAPDVS 120

Query: 135 EGKQLTSYKKAILNDSLKAEGIKLGDKVIDKSSSISLTVVGFVHNSMYGHGPVAFIDKD 194

EG+ L A++ ++SL+ +G+K+GD + K S L V GF ++ H P ++

Sbjct: 121 EGQTLRVAGGAVV-DESLREDGVKVGDLTLKPSGDQLRVSGFTR SARLNHQPGMYVSLA 179

Query: 195 IYTEINKKINPQYQFLPQALVMKNDKSISHLP-TQLEAVSKKDVIQHIPPYSAEQSTLNM 253

+ +K+NP+ A+ + + +L L ++ +Q +PGY EQ +L M

Sbjct: 180 RW----QKLNPRMHGTVNAVALPAAPQVNLGGADLSVTNRAQTLQVLPGYKEEQGSLTM 235

Query: 254 ILWVLVVASAGILGVFFYIITLQKRHEFSVMKAIGTKMSEIALFQLSQVIILALFGIIVG 313

I L+ +A +L FFY++TLQK +F ++KAIG +A ++Q++IL L + +

Sbjct: 236 IQVFLIAVAFAFVLATFFYVMTLQKTAQFGLLKAIGASNRTLGASVVAQMLITLTLAIAIA 295

Query: 314 DGLAVALSYVLPAQMPFVINWQNIILVSFVFLVIAMISSALSIVKVAKIDPV 365

+ + + +LPA MPF + NI S + LV+A ++S LS+ +VAK+DP+

Sbjct: 296 AAVTLGMVQLLPAGMPFHLTAANIASASGLLLVVAALASLLSVRRVAKVDPL 347

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6495> which encodes the amino acid sequence <SEQ ID 6496>. Analysis of this protein sequence reveals the following:

Possible site: 58

-2370-

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -12.31 Transmembrane 246 - 262 ( 233 - 270)  
 INTEGRAL Likelihood = -8.49 Transmembrane 327 - 343 ( 321 - 351)  
 INTEGRAL Likelihood = -1.01 Transmembrane 301 - 317 ( 301 - 317)

----- Final Results -----

bacterial membrane --- Certainty=0.5925(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF12525 GB:AE001863 hypothetical protein [Deinococcus radiodurans]  
 Identities = 101/360 (28%), Positives = 175/360 (48%), Gaps = 11/360 (3%)

Query: 1 MFLALNEMKQSKLRYGLIAGLLCLVAYLMFFLSGLAFGLMQENRSVAVDLWKADSVLLAKD 60  
 M+LAL E++ KLR LI G++ L+A+++F L+GL GL +++ S + A S + K+  
 Sbjct: 1 MYLALRELQHQKLRSLIGGIIVALIAFMVFMLTGLTRGLSRDSASLLDDTPAQSFVTTKE 60

Query: 61 ADATLTLSQVSRAQENQITADKVAPLAQLNTVAWSVKNPKDADKVKVSFLFGIDSNSFIRP 120  
 AD L S +S Q + + D A T K V L G+D F+ P  
 Sbjct: 61 ADGVLNRSFLSPEQVSALQQDNEDAAFAQTFFVSFSGDKQLSGV---LLGVDPGRGFLAP 117

Query: 121 NIVKGRFLKTNKEVVDQSLAKEEFAIGKDFYTSSSSQALTIVGYTQNFARFSVAPVVYM 180  
 ++ +G+ + V+D+SL +E+ +G S L + G+T++AR + P +Y+  
 Sbjct: 118 DVSEGTQLRVAGGAVVDESL-REDGVKVGDLTLKPSGDQLRVSGFTRSAARNHQPGRMYV 176

Query: 181 NLEAFETLKYGEPLPKDKQVVNAFITKGS--LTDYPKKDFQKLDIKTFITKLPGYSAQLL 238  
 +L ++ L P+ VNA + + D + + L PGY +  
 Sbjct: 177 SLARWQKLN-----PRMHGTVNAVALPAAPQVNLGGADLSVTNRAQTLQVLPGYKEEQG 231

Query: 239 TFGFMISFLVIISAIIGIFMYILTIQKAPIFGIMKAQGISNKTITTTAVLMQTFFLSFLG 298  
 + + FL+ ++A ++ F Y++T+QK FG++KA G SN+T+ +V+ Q L+ L  
 Sbjct: 232 SLTMIQVFLIAVAAFVLATFFYVMTLQKTAQFGLLKAIGASNRTLGSVVAQMLILTLLA 291

Query: 299 SGLGLLGTWLTSLLLPTVPFQSNWFLYLAIFVSMICFALLGTLFSVFNIIRIDPLKAIG 358  
 + T LLP +PF + ++ A L +L SV + ++DPL A+G  
 Sbjct: 292 VAIAAAVTLGMVQLLPAGMPFHLTAANIASASGLLLVVAALASLLSVRRVAKVDPLIALG 351

An alignment of the GAS and GBS proteins is shown below.

Identities = 96/356 (26%), Positives = 178/356 (49%), Gaps = 4/356 (1%)

Query: 15 AWKELTFYKKKYLLIELLIIVMMFMVFLSGLANGLGRAVSAAIENNPATYIILNEGAEQ 74  
 A E+ K +Y LI L+ ++ +++ FL SGLA GL + +A++ A + +L + A+  
 Sbjct: 4 ALNEMKQSKLRYGLIAGLLCLVAYLMFFLSGLAFGLMQENRSVAVDLWKADSVLLAKDADA 63

Query: 75 VITSSVLTTKDQTDNLNLDKSTLTNIQRSSSLTRQGHEKKIDISYFAIDKDSFMAPTLS 134  
 +T S ++ + + + + LN S+ K+ +S F ID +SF+ P +  
 Sbjct: 64 TLTL S QVSRAQENQITADKVAPLAQLNTVAWSVKNPKDADKVKVSFLFGIDSNSFIRPNIV 123

Query: 135 EGKQLTSYKKAIIILNDSLKAEGIKLGDVIDKSSSISLTVVGFVHNSMYGHGPVAFIDKD 194  
 +G+ + K+ ++ K E +G SSS +LT+VG+ N+ + PV +++ +  
 Sbjct: 124 KGRFLKTNKEVVDQSLAKEEFAIGKDFYTSSSSQALTIVGYTQNFARFSVAPVVYMNLE 183

Query: 195 IYTEIN-KKINPQYQFLPQALVMKNDKSISHLPTQ-LEAVSKKDVIQHPIGYSAEQSTLN 252  
 + + + P+ + + A + K S++ P + + + K I +PGYSA+ T  
 Sbjct: 184 AFETLKYGEPLPKDKQVVNAFITKG--SLTDYPKKDFQKLDIKTFITKLPGYSAQLLTFG 241

Query: 253 MILWVLVVASAGILGVFFYIITLQKRHEFSVMKAIGTKMSEIALFQLSQVIIILALFGIIV 312  
 ++ LV+ SA I+G+F YI+T+QK F +MKA G I L Q L+ G +  
 Sbjct: 242 FMISFLVIISAIIGIFMYILTIQKAPIFGIMKAQGISNKTITTTAVLMQTFFLSFLGSG 301

Query: 313 GDGLAVALSYVLPQMPFVINWQNIILVSFVFLVIAMISSALSIVKVAKIDPVEVI 368  
 G S +LP +PF NW + + + A++ + S+ + +IDP++ I  
 Sbjct: 302 GLLGTWLTSLLLPTVPFQSNWFLYLAIFVSMICFALLGTLFSVFNIIRIDPLKAI 357

-2371-

SEQ ID 8980 (GBS239) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 13; MW 64kDa).

GBS239-GST was purified as shown in Figure 227, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2100

A DNA sequence (GBSx2215) was identified in *S.agalactiae* <SEQ ID 6497> which encodes the amino acid sequence <SEQ ID 6498>. This protein is predicted to be heterocyst maturation protein (devA) (b0879). Analysis of this protein sequence reveals the following:

```

Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1751(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA05977 GB:AJ003195 ATP-binding subunit [Anabaena variabilis]
Identities = 87/225 (38%), Positives = 146/225 (64%), Gaps = 1/225 (0%)

Query: 3  AILELKHISKHYPDGDELLSILDNLDSVSAGEFVAILGPSGSGKSTLLSIAGLLLGADQ 62
      A++ +K ++ +Y G      IL +++L +  GE V + GPSGSGK+TLLS+ G L      +
Sbjct: 5  AVIAIKSLNHYYGKGALKRQILFDINLEIYPGEIVIMTGPSGSGKTTLLSLIGGLRSVQE 64

Query: 63  GSLYVNHENVTDLRSQRQRTQLRREALGFIFQSHQLLPYLTIQEQLEARFAKHVDKKT 122
      G+L      ++ SQ +  Q+RR ++G+IFQ+H LL +LT ++ +Q      +H ++ +
Sbjct: 65  GNLFGLVGLSGASQNKLVQIRR-SIGYIFQAHNLLGFLTARQNVQMAVELNEHISQEEA 123

Query: 123 LEEINKLLSDLGIEQCAHKYPNQLSGGQKQRAAIARAFINHPKVILADEPTASLDEERGR 182
      + +  +L +G+E      YP+ LSGGQKQR AIARA +N+P ++LADEPTA+LD++ GR
Sbjct: 124 IAKAEAMLKAVGLENRVDYYPDNLGGGQKQRVAIARALVNNPPLVLADEPTAALDKQSGR 183

Query: 183 QVTELIRQEVKSHNTAAIMVTHDERVLDLVDTVYRLKDGKLVKEN 227
      V E++++ K  T+ ++VTHD R+LD+ D +  ++DG L +++
Sbjct: 184 DVVEIMQLAKDQGTSLVLVTHDNRILDIADRVIMEDGILARDS 228

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6499> which encodes the amino acid sequence <SEQ ID 6500>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4181(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 103/224 (45%), Positives = 149/224 (65%), Gaps = 4/224 (1%)

Query: 3  AILELKHISKHYPDGDELLSILDNLDSVSAGEFVAILGPSGSGKSTLLSIAGLLLGADQ 62
      ++L K ++K + DG  ++ L  D S+ AGEFVAI+GPSGSGKST L+IAG L
Sbjct: 3  SVLTFKQVTKTFQDGHHEINALKATDFSIRAGEFVAIIGPSGSGKSTFLTLAGGLQTPSS 62

Query: 63  GSLYVNHENVTDLRSQRQRTQLRREALGFIFQSHQLLPYLTIQEQLEARFAKHVDKKT 122

```

-2372-

G L ++ + T LS+++R++LR +++GFI Q+ L+P+ T+Q+QL+ H  
 Sbjct: 63 GQLIIDGTDYTHLSEKERSRLRFKSVGFILQASNLIPFSTVQQQLE----LVDHLTGSKE 118

Query: 123 LEEINKLLSDLGIEQCAHKYPNQLSGGQKQRAAIAARAFINHPKVILADEPTASLDEERGR 182  
 + N+L DLGI H+ P +LSGG++QRAAIAARA + P +ILADEPTASLD E+  
 Sbjct: 119 KAKANQLFDDLGITGLKHQLPQELSGGERQRAAIAARALYHDPALILADEPTASLDTEKAY 178

Query: 183 QVTELIRQEVKSHNTAAIMVTHDERVLDLVDVTYRLKDGKLVKE 226  
 +V +L+ +E K N A IMVTHD+R+L D VYR++DG+L +E  
 Sbjct: 179 EVVKLLAKESKEKNKAIIMVTHDDRMLKYCDKVYRMQDGELOCQ 222

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2101

- 15 A DNA sequence (GBSx2216) was identified in *S.agalactiae* <SEQ ID 6501> which encodes the amino acid sequence <SEQ ID 6502>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2645(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB64972 GB:AJ012050 VicR protein [Enterococcus faecalis]  
 Identities = 86/229 (37%), Positives = 132/229 (57%), Gaps = 10/229 (4%)

Query: 3 KILVVEDNIVQQKIITTKLTQEGYQFITASNGQEALNCLDTEEVQLIITDIMPMMDGYQ 62  
 KILVV+D +I+ L +EGY+ TA +G+EAL ++ E LII D+M+P MDG +  
 Sbjct: 52 KILVVDEKPISEIVKYNLVKEGYEVFTAYDGEEALEKVEEVEPDLIILDMLPKMDGLE 111

Query: 63 LIQELRSAAYNPPIIVMTAKSQMEDMTKGFLGADDYMKPVQLQELALRIKALLRR--- 119  
 + +E+R +++PII++TAK D G LGADDY+ KP +EL R+KA LRR  
 Sbjct: 112 VAREVRK-THDMPPIIMVTAKDSEIDKVLGLELGADDYVTKPFSNRELVARVKANLRRGAT 170

Query: 120 ----ANIVAQHQLIIGNTCLNEDELSLKYFEQEIIFPQKEFRVLFHLLSYPNRIFTRLEL 175  
 A + Q +L IG+ ++ D + ++I +EF +L++L + ++ TR L  
 Sbjct: 171 NAKEAEVTTQSELTIGDLTIHPDAYMVSKRGEKIELTHREFELLYLAKHIGQVMTREHL 230

Query: 176 LDSIWGMDTDLDERVVDACINKIRKVEHLPDFK--IETVRGVGYRAKN 222  
 L ++WG D D R VD + ++R K+E P + T RGVGY +N  
 Sbjct: 231 LQTVWGYDYFGDVRTVDVTVRRLREKIEDSPSHPTYLVTRGVGYYLNRN 279

- 45 There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2102

- 50 A DNA sequence (GBSx2217) was identified in *S.agalactiae* <SEQ ID 6503> which encodes the amino acid sequence <SEQ ID 6504>. This protein is predicted to be sensor protein. Analysis of this protein sequence reveals the following:

Possible site: 38  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -8.97 Transmembrane 53 - 69 ( 47 - 77)

-2373-

## ----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC62214 GB:AF049873 sensor protein [Lactococcus lactis]  
 Identities = 97/307 (31%), Positives = 169/307 (54%), Gaps = 16/307 (5%)

Query: 57 SALAVVFLSLVIASISMWYGSYHLTKPILDISHIVSNVADGDFEGHIYRNSNRKSYEYY 116  
 + LAV+ +L++ + S++Y + +T+P+L I +A GD + N+  
 Sbjct: 170 AVLAVI--TLIVTAFSIFYITRTVTRPLLKIKLGTDKIAQGDLSIQLNVNTE----- 219

Query: 117 NELDELSINQMIVSLSHMDHMRKDFITNVSHLKTPIAAVANIVELLQDPELDEETQS 176  
 +EL EL++SI + L M R +F+++V+HEL+TP+ + ++ E ++  
 Sbjct: 220 DELGELAKSIEDLAEKLDPMKRENEFLSSVAHELRTPLTFIKGYADIANRSTTSLEDKT 279

Query: 177 ELLGLVKTESLRLTRLCDTMLQMSRVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRI 236  
 + L +++ ES LT+L + ++ +++++ E V + E I + + ++ + KRI  
 Sbjct: 280 QYLRIIREESRHLTQLMEDLMNLAQLENGFKVEKHQVLIQELINEVSVKSGVFSEKRI 339

Query: 237 NFQLDSKPYPYVYSNSDLLM--QVINLLDNAIKYSEDIVDLSVRMEETNNHYLRVIISDK 294  
 NF L S Y+N D + QV +NLL NA KYS D D+ + ++ +++ISDK  
 Sbjct: 340 NF-LISGEGNFYANIDFMRIEQVLVNLMLNAYKYSADESDIKLAFIPEKENF-KIVISDK 397

Query: 295 GRGISQYDVQHIFDKFYQADQSHNQ--GNGLGLAIVKRIIVLCGRISVSSQLEIGTEF 352  
 G GI + D+ +IF++FY+ D+S + G GLGLAIV+ I+ G+I V S GT F  
 Sbjct: 398 GEGIQEQDLPYIFERFYRVDKSRTRTTGGVGLGLAIVQDIVKKHNGKIIVESIQNQGTTF 457

Query: 353 CVELPLS 359  
 +ELP S  
 Sbjct: 458 IIELPYS 464

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8981> and protein <SEQ ID 8982> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10  
 McG: Discrim Score: 4.84  
 GvH: Signal Score (-7.5): 0.179999  
 Possible site: 35  
 >>> Seems to have a cleavable N-term signal seq.  
 ALOM program count: 1 value: -8.97 threshold: 0.0  
 INTEGRAL Likelihood = -8.97 Transmembrane 50 - 66 ( 47 - 77)  
 PERIPHERAL Likelihood = 1.27 324  
 modified ALOM score: 2.29

\*\*\* Reasoning Step: 3

## ----- Final Results -----

bacterial membrane --- Certainty=0.4588(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

31.9/57.3% over 293aa

Lactococcus lactis

GP|3687664| sensor protein Insert characterized

ORF01881(478 - 1377 of 1677)

GP|3687664|gb|AAC62214.1||AF049873(171 - 464 of 464) sensor protein [Lactococcus lactis]  
 %Match = 12.9

-2374-

%Identity = 31.9 %Similarity = 57.3

Matches = 94 Mismatches = 121 Conservative Sub.s = 75

```

5      339      369      399      429      459      489      519      549
MTKLRRFRFPLRFYFTLMFVLTMFLSVLASLLLVAAIVFTFFQGVLTTHVLQVSALAVVFLSLVIASISMWYGSYHLTKP
| :: : : : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
EKNKKESLHFWLGDKYIVSKSRIQSNQKIVGSVYMFSTRPIQKMFNFTGIFAVLAVITLIVTAFSIFYITRIVTRP
      130      140      150      160      170      180      190

10     579      609      639      669      699      729      759      789
ILDISHIVSNVADGDFEGHIYRNSNRKRSY EYYNELDELSE SINQMIVSLSHMDHMRKDFITNVSHELKTPIAAVANIVE
: | | : | | : : : | : : | | : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
LLKIKLGTDKIAQGDLSIQNLVNTE-----DELGELAKSIEDLAEKLD FMKRENEFLSSVAHELRTPLTFIKGYAD
      210      220      230      240      250      260

15     819      849      879      909      939      969      999      1029
LLQDPELDEETQSELLGLVKTESLRRLTRLCDTMLQMSRVDNQETIGELSSVRVDEQIRQAMISLTERWQAKRINFQLD SK
: | : : | : : | | : | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
IANRSTTSLEDKTOYLRIIREESRHLTQLMEDLMNLAQLEENGFKVEKHQVLIQELINEVSVKSGVGFSEKRINF-LISG
      280      290      300      310      320      330      340

20     1059     1083     1113     1143     1173     1203     1233
PYTVYSNSDLL--MQVWINLLDNAIKYSEDIVDLSVRMEETNNHYLRVVISDKGRGISQYDVQHIFDKFYQADQSHNQ-
| : | : : | : | | | | | | : : : : : : | | : | : | : : | : | : : : | : | :
EGNFYANIDFMRIEQVLVNLMLNAYKYSADESDIKLAFIPEKENF-KIVISDKGEGIPEQDLPYIFERFYRVDKSRTRTT
      360      370      380      390      400      410      420

25     1287     1317     1347     1377     1407     1437     1467     1497
-GNGLGLAIVKRIIVLCKGRISVSSQLEIGTEFCVELPLS*LFKTTITANWQLLFYLFNRNKYTKNRQKL*KYLTINIASV*
| | | | | | : | : | : | | | | : | | : | | : | : | : | : | : | : | : | : | : | : | :
GGVGLGLAIVQDIVKKHNGKIIVESIQNQGTTFIIELPYS
      440      450      460

```

SEQ ID 8982 (GBS170d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 181 (lane 4; MW 35kDa) and in Figure 123 (lane 5-7; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 123 (lane 2-4; MW 60kDa) and in Figure 184 (lane 3; MW 60kDa). Purified GBS170d-GST is shown in Figure 243, lane 7; purified GBS170d-His is shown in Figure 234, lanes 5-6.

### Example 2103

A DNA sequence (GBSx2218) was identified in *S.agalactiae* <SEQ ID 6505> which encodes the amino acid sequence <SEQ ID 6506>. Analysis of this protein sequence reveals the following:

Possible site: 21

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.0502(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06906 GB:AP001518 argininosuccinate synthase
(citrulline-asparate ligase) [Bacillus halodurans]
Identities = 262/396 (66%), Positives = 321/396 (80%), Gaps = 1/396 (0%)

Query: 1 MGKEKLILAYSGGLDTSVAIAWLK-KDYDVIAVCMVDVGEKGLDLDFIHDKALTIGAIESYI 59
M K+K++LAYSGLDTSVAI WL K YDVIAV +DVGEKGLD+F+ +KAL +GAIESY
Sbjct: 1 MSKKKVVLAYSGGLDTSVAIKWLSDKGYDVIAVGLDVGEKGLDFVKEKALKVGAIESYT 60

Query: 60 LDVKDEF AEHFVLPALQAHAMYEQKYPLVSALS SRPIIAQKLVEMAHQTGATTIAHGCTGK 119
+D K EF AE FVLPALQAH+YEQKYPLVSALS SRP+I++KLVE+A QTGA +AHGCTGK
Sbjct: 61 IDAKKEFAEEFVLPALQAHALYEQKYPLVSALS SRPLISKLVETAEQTGAQAVAHGCTGK 120

```



-2375-

5 Query: 120 GNDQVRFEVAIAALDPELKVIAPVREWKWHREEEITFAKANGVPIPADLDNFPYSIDQNLW 179  
 GNDQVRFEV+I AL+P L+V+APVREW W R+EEI +AK N +PIP DLDNPYS+DQNLW  
 Sbjct: 121 GNDQVRFEVSIQALNPNLEVLAPVREWAWSRDEEIEYAKKNNIPIDLDNFPYSVDQNLW 180

10 Query: 180 GRANECGVLENPNWQAPEEAFGITKSPEEAPDCAEYIDITFQNGKPIAINNQEMTLADLI 239  
 GR+NECG+LE+PW PE A+ +T + E+APD E ++I F+ G P+ +N + + +LI  
 Sbjct: 181 GRSNECGILEDPWATPPEGAYELTVAIEDAPDQPEIVEIGFEKGIPVTLNGKSYPVHELI 240

15 Query: 240 LSLNEIAGKHGIGRIDHVENRLVGIKSREIYECPAAMVLLAAHKEIEDLTLVREVSHFKP 299  
 L LN+IAGKHG+GRIDHVENRLVGIKSRE+YEC P AM L+ AHKE+EDLTL +EV+HFKP  
 Sbjct: 241 LEINQIAGKHGIGRIDHVENRLVGIKSREYECPGAMTLIKAHKEIEDLTLTKEVAHFKP 300

20 Query: 300 ILENELSNLIYNALWFSPTAKIAYVKETQKVNGTTKVLYKGSQVVARHSSNSLYD 359  
 ++E +++ LIY LWFSP A+ A++KETQ V G +VKL+KG A V R S SLY+  
 Sbjct: 301 VVEKKIAELIYEGLWFSPLQPALSAFLKETQSTVTGVVRVKLFKGHAIVEGRKSEYSLYN 360

Query: 360 ENLATYTAADSFDQDAAVGFIKLWGLPTQVNAQVVK 395  
 E LATYT D FD +AAVGFI LWGLPT+V + VNK  
 Sbjct: 361 EKLATYTPDDEFDHNAAVGFISLWGLPTKVYSMVNK 396

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 25 Example 2104

A DNA sequence (GBSx2219) was identified in *S.agalactiae* <SEQ ID 6507> which encodes the amino acid sequence <SEQ ID 6508>. This protein is predicted to be argininosuccinate lyase (argH). Analysis of this protein sequence reveals the following:

30 Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2131(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB06905 GB:AP001518 argininosuccinate lyase [Bacillus halodurans]  
 Identities = 284/454 (62%), Positives = 350/454 (76%)

Query: 6 KLWGGRFESSLEKWVEFGASISFDQKLAPYDMKASMAHVMTMLGKTDIISQEEAGLIKDG 65  
 KLWGGRF + E WV+EFGASI FDQ+L D++ S+AHVTML K+ I++ EE IK G  
 Sbjct: 3 KLWGGRFTKTAEAWVDEFGASIGFDQQLVEEDIEGSLAHVTMLEKSGILANEEVEQIKKG 62

45 Query: 66 LKILQDKYRAGQLTFSISNEDIHMNIESLLTAEIGE VAGKLHTARSRNDQVATDMHLYLK 125  
 L IL +K + G+L +S+++NEDIH+NIE LL EIG V GKLHT RSRNDQVATDMHLYL+  
 Sbjct: 63 LHILLEKAKKGELNYSVANEDIHLNIEKLLIDEIGPVGGKLHTGRSRNDQVATDMHLYLR 122

50 Query: 126 DKLQEMMKLLHLRTTLVNLAENHIYTVMPGYTHLQHAQPISFGHLMAYYNMFTRDTER 185  
 + +E+++ + +++ LV A+ H+ T++PGYTHLQ AQPI SF HHL+AY+ M RD R  
 Sbjct: 123 KQTKELQLVKNVQAALVEQAKQHVETLIPGYTHLQRAQPISFAHLLAYFWMLERDYGR 182

55 Query: 186 LEFNMKHTNLSPLGAAALAGTTFPIDRHMTTRLLDFEKPYSNSLDAVSDRDFIIEFLSNA 245  
 E ++K N+SPLGA ALAGTTFPIDR T LL F+ Y NSLDAVSDRDFI+EFLS +  
 Sbjct: 183 YEDSLKRLNVSPLGAGALAGTTFPIDREYTAELLGFDGIYENSLDAVSDRDFIVEFLSAS 242

Query: 246 SILMMHLSRFCEEIINWCSYEQFITLSDTFSTGSSIMPQKKNPDMAE LIRGKTGRVYGN 305  
 S+LM HLSR CEE+I W S E+QF+ + D F+TGSSIMPQKKNPDMAE LIRGKTGRVYG+  
 Sbjct: 243 SLLMTHLSRLCEELILWSSQEFQFVEMDDAFATGSSIMPQKKNPDMAE LIRGKTGRVYGS 302

60

-2376-

Query: 306 LFSLLTVMKSLPLAYNKDLQEDKEGMFDSVETVSAIEIMANMLETMTVNEHIMMTSTET 365  
 LFSLLTV+K LPLAYNKD+QEDKEGMFD+V+TV ++ I A M++TM V E M +  
 Sbjct: 303 LFSLLTVLKGLPLAYNKDMQEDKEGMFDVKT VKGSLAIFAGMIQTMKVKEETMTKAVHQ 362

5 Query: 366 DFSNATELADYLASKGVFPFRKAHEIVGKLVLECSKNGSYLQDIPKYYQEISELIENDIY 425  
 DFSNATELADYLA+KG+PFR+AHE+VGKLV L C + G YL D+PL Y+ S+L + DIY  
 Sbjct: 363 DFSNATELADYLATKGMPFREAHEVVGKLVLLCIQKGIYLLDLPLSDYKAASDLFDEDIY 422

10 Query: 426 EILTAKTAVKRRNSLGGTGFQVKKQILLARKEL 459  
 ++L KT V RR S GGTGF +VKK I A K L  
 Sbjct: 423 DVLQPKTVVARRTSAGGTGFTEVKKAIKAEKIL 456

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 15 Example 2105

A DNA sequence (GBSx2220) was identified in *S.agalactiae* <SEQ ID 6509> which encodes the amino acid sequence <SEQ ID 6510>. This protein is predicted to be class-II aldolase (fba). Analysis of this protein sequence reveals the following:

20 Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9289> which encodes amino acid sequence <SEQ ID 9290> was also identified. Analysis of this sequence reveals:

30 GvH: Signal Score (-7.5): -2.92  
 Possible site: 42  
 >>> Seems to have no N-terminal signal seq.  
 ALOM program count: 0 value: 0.37 threshold: 0.0  
 PERIPHERAL Likelihood = 0.37 66  
 35 modified ALOM score: -0.57

\*\*\* Reasoning Step: 3

40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA16889 GB:AB050113 class-II aldolase [Streptococcus bovis]  
 Identities = 221/242 (91%), Positives = 234/242 (96%)

50 Query: 1 MAIVSAEKVFQAAARDNGYAVGGFNTNNLEWTQAILRAAEAKKAPVLIQTSMGAAKYMGGY 60  
 MAIVSAEKF++AAR+NGYAVGGFNTNNLEWTQAILRAAEAKKAP+LIQTSMGAAKYMGGY  
 Sbjct: 1 MAIVSAEKFIKAARENGYAVGGFNTNNLEWTQAILRAAEAKKAPILIQTSMGAAKYMGGY 60

Query: 61 KLCKQLIETLVESMGITVPVAIHLHDHGHYDDALECIEVGYTSIMFDGSHLPVEENLEKAR 120  
 KLCK LIE LVESMGITVPVAIHLHDHGH++DALECIEVGYTS+MFDGSHLPVEENLEKA+  
 55 Sbjct: 61 KLCKTLIENLVESMGITVPVAIHLHDHGHFEDALECIEVGYTSVMFDGSHLPVEENLEKAK 120

Query: 121 EVVAKAHAKGISVEAEVGTIGGEEDGIVGKELAPIEDAKAMVETGIDFLAAGIGNIHGP 180  
 EVVAKAHAKG+SVEAEVGTIGGEEDGIVG KELAPIEDAKAMV TGIDFLAAGIGNIHGP  
 Sbjct: 121 EVVAKAHAKGVSVEAEVGTIGGEEDGIVGGELAPIEDAKAMVATGIDFLAAGIGNIHGP 180

-2377-

Query: 181 YPANWEGLDLDHLKKLITEAVPGFPFIVLHGGSGIPDDQIQEAIKLGVAKVNVNTECQLAFC 240  
 YPANW+GL LDHLKKLT AVPGFPFIVLHGGSGIPDDQI+ AIKLGVAKVNVNTECQ+AF  
 Sbjct: 181 YPANWQGLHLDHLKKLTAAVPGFPFIVLHGGSGIPDDQIKAAIKLGVAKVNVNTECQIAFA 240

5 Query: 241 QA 242  
 +A  
 Sbjct: 241 KA 242

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6511> which encodes the amino acid sequence <SEQ ID 6512>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2930(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 217/242 (89%), Positives = 228/242 (93%)

Query: 1 MAIVSAEKFFVQAARDNGYAVGGFNTNNLEWTQAILRAAEAKAPVLIQTSMGAAKYMGGY 60  
 MAIVSAEKFFVQAAR+NGYAVGGFNTNNLEWTQAILRAAEAK+APVLIQTSMGAAKYMGGY  
 25 Sbjct: 1 MAIVSAEKFFVQAARENGYAVGGFNTNNLEWTQAILRAAEAKQAPVLIQTSMGAAKYMGGY 60

Query: 61 KLCKQLIETLVESMGITVPAIHLHDHGHYDDALECIEVGYSIMFDGSHLPVEENLEKAR 120  
 K+C+ LI LVESMGITVPAIHLHDHGHY+DALECIEVGYSIMFDGSHLPVEENL K  
 30 Sbjct: 61 KVCQSLITNLVESMGITVPAIHLHDHGHYEDALECIEVGYSIMFDGSHLPVEENLAKTA 120

Query: 121 EVVAKAHAKGISVEAEVGTIGGEEDGIVKGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180  
 EVV AHAKG+SVEAEVGTIGGEEDGI+GKGELAPIEDAKAMVETGIDFLAAGIGNIHGP  
 Sbjct: 121 EVVKIAHAKGVSVEAEVGTIGGEEDGIIGKGELAPIEDAKAMVETGIDFLAAGIGNIHGP 180

35 Query: 181 YPANWEGLDLDHLKKLITEAVPGFPFIVLHGGSGIPDDQIQEAIKLGVAKVNVNTECQLAFC 240  
 YP NWEGL LDHL+KLT AVPGFPFIVLHGGSGIPDDQI+EAI+LGVAKVNVNTE Q+AF  
 Sbjct: 181 YPENWEGLLDHLLEKLTAAVPGFPFIVLHGGSGIPDDQIKEAIRLGVAKVNVNTEQIAFS 240

40 Query: 241 QA 242  
 A  
 Sbjct: 241 NA 242

SEQ ID 9290 (GBS683) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 8 & 10; MW 55kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 150 (lane 11-13; MW 30kDa) and in Figure 184 (lane 11; MW 30kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2106

- 50 A DNA sequence (GBSx2221) was identified in *S.agalactiae* <SEQ ID 6513> which encodes the amino acid sequence <SEQ ID 6514>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2775(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2378-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAA88585 GB:M18954 unknown protein [Streptococcus mutans]  
Identities = 109/229 (47%), Positives = 156/229 (67%), Gaps = 1/229 (0%)

Query: 1 MFSGKRLKKRRITLGYSQSELADKLHINRSSYFNWENEKTKPNQSNLKQLAILLDVPETY 60  
MFS ++LK+RR LG SQ++ ADKL I+R SYFNWE KTKPNQ NL +LA LL V Y  
10 Sbjct: 1 MFSSQKLKERRKKLGLSQAQTADKLGISRPSYFNWEIGKTKPNQKNLDKLAHLKVD SAY 60

Query: 61 FESEYKIVNTYLQSLQSQEYKVEKYAEELLQTQKVHEKIVPLFAVEVLSEIQLSAGPGEG 120  
F S++ IV Y +L+ N+ K KY++ LL+ Q ++ +LSAG G  
Sbjct: 61 FLSQHDIVEIYTRLNESNKTTLKYSQHLLQDQKKRNLMMKNKRYPYRVYEKLSAGTGYS 120

15 Query: 121 LYDEFETETVYSEDEYTGFDIATWISGNSMEPVYKDGEVALIRSTGFDHDGAVYALNWNG 180  
+ + +TV+ ++E D A+WI G+SMEP++ +GEVALI+ TGFD+DGA+YA++W+G  
Sbjct: 121 YFGDGNFDTVFYDEEID-HDFASWIFGDSMEPIFLNGEVALIKQTGFDYDGAIFYAIDWDG 179

20 Query: 181 SLYIKKLYREEDGFRMVSINPDVAERFIPFEDEIRIVGKIVGHFMPVIG 229  
YIKK+YREE G R+VS+N A++F P+++ RI+G IVG+F+P+ G  
Sbjct: 180 QTYIKKVYREETGLRLVSLNKYADKFAPYDENPRIIGLIVGNFIPLEG 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6515> which encodes the amino acid sequence <SEQ ID 6516>. Analysis of this protein sequence reveals the following:

25 Possible site: 38  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
30 bacterial cytoplasm --- Certainty=0.4340 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 84/209 (40%), Positives = 130/209 (62%), Gaps = 9/209 (4%)

35 Query: 25 LHINRSSYFNWENEKTKPNQSNLKQLAILLDVPETYFESEYKIVNTYLQSLQSQEYKVEK 84  
LH+N+ + NWE K PN+ +L L L +V YF+ Y+++ Y QL++ N+EKV  
Sbjct: 5 LHVNMKTISNWEKGKNIPNEKHLNALHLFNVTSDYFDPNYRLLTYPNQLTISNKEKVIG 64

40 Query: 85 YAEELLQTQ-----KVHEKIVPLFAVEVLSEIQLSAGPGEGLYDEFETETVYSEDEYTG 138  
Y+E LL Q + +K L+A V LSAG G + + + V+ DE  
Sbjct: 65 YSERLLNHQIDKSKDLIDKPSQLYAYRVYES--LSAGTGYSYFGDGNFDDVVFY-DEQLE 121

45 Query: 139 FDIATWISGNSMEPVYKDGEVALIRSTGFDHDGAVYALNWNGSLYIKKLYREEDGFRMVS 198  
+D A+W+ G+SMEP Y +GEV LI+ FD+DGA+YA+ W+G YIKK++RE++G R+VS  
Sbjct: 122 YDFASWVFGDSMEPTYLNGEVVLKQNSFDYDGAIFYAVEWDGQTYIKKVFREDEGLRLVS 181

Query: 199 INPDVAERFIPFEDEIRIVGKIVGHFMPV 227  
+N +++F P+ +E RI+GKI+ +F P+  
50 Sbjct: 182 LNKKYSDKFAPYSEEPRIIGKIIANFRPL 210

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2107

55 A DNA sequence (GBSx2222) was identified in *S.agalactiae* <SEQ ID 6517> which encodes the amino acid sequence <SEQ ID 6518>. Analysis of this protein sequence reveals the following:

Possible site: 41  
>>> Seems to have no N-terminal signal sequence

-2379-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2387(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 10 Example 2108

A DNA sequence (GBSx2223) was identified in *S.agalactiae* <SEQ ID 6519> which encodes the amino acid sequence <SEQ ID 6520>. This protein is predicted to be UmuC MucB homolog (uvrX). Analysis of this protein sequence reveals the following:

Possible site: 47

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2195(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9925> which encodes amino acid sequence <SEQ ID 9926> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAC98439 GB:L29324 UmuC MucB homolog [Streptococcus pneumoniae]  
 Identities = 303/436 (69%), Positives = 360/436 (82%)

Query: 39 LHTSLCVMSRADNSAGLILASSPMFKKVFQKGNVGRAYDLFPDVHTRKFNYRAKISGLP 98  
 L LCVMSRADNSAGLILASSPMFKKVFQK NVGR+YDLFPDV TRKF+YY AK GLP

30 Sbjct: 5 LRLRLCVMSRADNSAGLILASSPMFKKVFQKSNVGRSYDLFPDVKTRKFSYNAKKQGLP 64

Query: 99 TDAKFVFSFIENWAKRTFIVPPRMDLYIQNLEIQKVFQNYADPTDILPYSIDEGFIDLTS 158  
 T +V +IE WAK T IVP L I N+EIQK+FQ++A P DI PYSIDEGFIDLTS

35 Sbjct: 65 TTIDYVRYIEEWAKSTVIVPREWILTIAVMEIQKIFQDFAAPDDIYPYSIDEGFIDLTS 124

Query: 159 SLNYFVEDKSLSRDKLDVVSIAKIQHDIWEKTGVYSTVGMSNANPLLAKLALDNEAKTTA 218  
 SLNYFV DKS+SRDKLD++SA IQ IW KTG+YSTVGMSNANPLLAKLALDNEAK T

Sbjct: 125 SLNYFVPDKSISRDKLDIISAAIQKKIWRKTGIYSTVGMSNANPLLAKLALDNEAKKTP 184

40 Query: 219 TMRANWSYEDVETKVWNIPKMTDFWGIGSRTEKRLNKLGIYSIKELANCDPTILKKEFGV 278  
 TMRANWSYEDVE KVV IPKMTDFWGIG+R EKRL+ LGI+SIKELA +P ++KKE G+

Sbjct: 185 TMRANWSYEDVEKKVWTIPKMTDFWGIGNRMEKRLHNLGIFSIKELAQANPDLIKELGI 244

45 Query: 279 IGVQHFHANGIDESNVHEPYRPKAVGIGNSQVLHKDYTRQSDIELVLEMAEQVAIRLR 338  
 +G++ WFHANGIDESNVH+PY+PK+ GIGNSQVL KDY +Q DIE++LREMAEQVA+RLR

Sbjct: 245 MGLELWFHANGIDESNVHKPKPKSGIGNSQVLPKDYIKQDIEIILREMAEQVAVRLR 304

Query: 339 RRHKATVVAINVGYSNFENKKSINVQRKINPNNRTLQFQDEVVSLFRSKYDGGAVRSIA 398

50 Sbjct: 305 RSGKKATVVSIIHLGYSKVEQKRSINTQMKIEPTNQTALLTNYVLKLFHTKYTSGAIRNVA 364

Query: 399 VRYDGLVDENFAVISLFDFFESEKEEKLTTIDSIRDRFGFLAVQKASSILENSRAISR 458  
 V Y GLVDE+F +ISLFD E+ EKEE+L++ ID+IR FGF ++ K ++L + SR I+R

Sbjct: 365 VNYSGLVDESFGILSLFDIEKIEKEERLQSAIDAIRTEFGFTSLKGNALDQASRTIAR 424

55

Query: 459 SRLVGGHSAGGLEGLK 474  
 S+L+GGHSAGGL+GLK

-2380-

Sbjct: 425 SKLIGGHSAGGLDGLK 440

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 5 Example 2109

A DNA sequence (GBSx2224) was identified in *S.agalactiae* <SEQ ID 6521> which encodes the amino acid sequence <SEQ ID 6522>. Analysis of this protein sequence reveals the following:

Possible site: 45

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4016(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 20 Example 2110

A DNA sequence (GBSx2225) was identified in *S.agalactiae* <SEQ ID 6523> which encodes the amino acid sequence <SEQ ID 6524>. Analysis of this protein sequence reveals the following:

Possible site: 32

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2088(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:AAG13001 GB:AF227520 unknown [Streptococcus pneumoniae]

Identities = 68/122 (55%), Positives = 89/122 (72%), Gaps = 6/122 (4%)

Query: 1 MIDRSYLPFKVAREYQDRKMAKWMGFFLSEHTAGLDSELNKVDYTSSELSISDKLLLLNQL 60

MIDRSYLPF+ AREYQD KM KWMGFFLSEHT+ L + NKV Y S+LS+ KLLLL+Q+

Sbjct: 1 MIDRSYLPFQSAREYQDTKMQKWMGFFLSEHTSALTDDANKVTYMSDLSLEKKLLLLLSQV 60

Query: 61 YSNQLNGIIVPGQ---YYSGKVDNLTFFNHVSLKTKTGFSIPIKDILSIDL--EVEYE 114

Y+ QLN I V + Y+G + +LT + + +KT TG +++ +KDI+SI+L EV YE

Sbjct: 61 YAGQLNTRIHVVKNNQVSYTGTIPSLTKDFILIKTTTGHINLKLKDIVSIELVEEVLYE 120

Query: 115 SA 116

SA

Sbjct: 121 SA 122

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2111**

A DNA sequence (GBSx2226) was identified in *S.agalactiae* <SEQ ID 6525> which encodes the amino acid sequence <SEQ ID 6526>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4025(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9927> which encodes amino acid sequence <SEQ ID 9928> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2112**

A DNA sequence (GBSx2227) was identified in *S.agalactiae* <SEQ ID 6527> which encodes the amino acid sequence <SEQ ID 6528>. This protein is predicted to be soluble transducer HtrXIII. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5246(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2113**

A DNA sequence (GBSx2228) was identified in *S.agalactiae* <SEQ ID 6529> which encodes the amino acid sequence <SEQ ID 6530>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.5131(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-2382-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2114**

A DNA sequence (GBSx2229) was identified in *S.agalactiae* <SEQ ID 6531> which encodes the amino acid sequence <SEQ ID 6532>. This protein is predicted to be pXO2-78. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2105(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF13682 GB:AF188935 pXO2-78 [Bacillus anthracis]

Identities = 101/314 (32%), Positives = 147/314 (46%), Gaps = 46/314 (14%)

Query: 27 SGQIYEHPDHDSFRIFADTNTFFKWFSDIQGDVIDFVQLVAGVSFKKALSYLETG--GFE 84

S + Y +HDS I N F W SR + G++I FVQ V SF A+ L G +E

Sbjct: 39 SERYYRLTEHDSLII DRKKNQFYWNSRGVNGNIIKFVQEVEDASFPGAMQRLLDGEQDYE 98

Query: 85 EAKVIEETYQPFQYYLREEP----FQQARTYLKDIRGLSNQTINSFGRQGLLAQATYQAE 140

+A I +P+ Y E+ F +AR YL + R + Q +++ +GL+ Q Y

Sbjct: 99 KASEITFVSEPYDYEHEFEQKEVSRFDRAREYLI EERKIDPQVVDALHNKGLIKQDKYN-- 156

Query: 141 SVLVFKSFDHNGTILQAASLQGLVKNEEKYDRGYLKKIMKSGSHGVGISFDIGNPKRLIFC 200

+VL G + S QG+VK++ KY RG K I K S + G + G P+ L F

Sbjct: 157 NVLFLWKDRETGAVMGGSEQGVVKS D-KYKRGAWKSIQKNSTANYGFNVLNGEPRNLKFY 215

Query: 201 ESVIDMMSYYQLHQQLSDVRLISMEGLKLSVIAYQTLRLAAEEQGKLAFLDTVKPIRLS 260

ES ID++SY LH+ L D LISMEGLK VI +

Sbjct: 216 ESDIDL LSYATLHKHNLK DTHLISMEGLKPQVI-----FN 250

Query: 261 HYLQAIQETTTFFQTHSNVITMAVDNDEAGREFYQKL-----SDKGFPIFQ-DLPPLQ 312

+Y++A + + +++ VDND+AG+ F ++L +D F+ + P

Sbjct: 251 YMKACERIGDV----PDSL SLCVDNDKAGAFVERLIHFRYEKNDGSIVAFKPEYPQAP 306

Query: 313 RLETKSDWN DIVKR 326

E K DWND KR

Sbjct: 307 SEEKKWDWNDECKR 320

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2115**

A DNA sequence (GBSx2230) was identified in *S.agalactiae* <SEQ ID 6533> which encodes the amino acid sequence <SEQ ID 6534>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.7013(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>



The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2116

A DNA sequence (GBSx2231) was identified in *S.agalactiae* <SEQ ID 6535> which encodes the amino acid sequence <SEQ ID 6536>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2117

A DNA sequence (GBSx2232) was identified in *S.agalactiae* <SEQ ID 6537> which encodes the amino acid sequence <SEQ ID 6538>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.6726(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9373> which encodes amino acid sequence <SEQ ID 9374> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2118

A DNA sequence (GBSx2233) was identified in *S.agalactiae* <SEQ ID 6539> which encodes the amino acid sequence <SEQ ID 6540>. This protein is predicted to be phosphoglucomutase (manB). Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

-2384-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2147(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9355> which encodes amino acid sequence <SEQ ID 9356> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB96418 GB:AJ243290 phosphoglucomutase [Streptococcus thermophilus]  
Identities = 391/465 (84%), Positives = 424/465 (91%), Gaps = 1/465 (0%)

Query: 1 MAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA 60  
+A HGIKSYVFE+LRPTPELSFAVRHL+ +AGIM+TASHNPAPFNGYKVYQ+DGGQ+PPA

15 Sbjct: 107 LAAHGIKSYVFESLRPTPELSFAVRHLHTFAGIMTASHNPAPFNGYKVYGEDGGQMPPA 166

Query: 61 DADALTD FIRAIENPF FAVELADLDESKSSGLIQVIGEDVDIEYLREV KDVNINQDLINNF 120  
DADALTD+IRAI+NPF V+LADL++SK+SGLI++IGE+VD EYL+EVKDVNINQDLIN +

20 Sbjct: 167 DADALTDYIRAI DNPF TVKLADLDESKASGLIETIGENVDAEY LKEVKDVNINQDLIN EY 226

Query: 121 GKDMKIVYTP LHGTGEMLTRRALAQAGFESV VVVSQAKADPDFSTVKSPNPESQA AFAL 180  
G+DMKIVYT LHGTGEML RRALAQAGF++V VVE+QA DF TVKSPNPE+Q AFAL

Sbjct: 227 GRDMKIVYTS LHGTGEML VRRALAQAGF DAVQVVEAQAVPHADFLTVKSPNPENQDAFAL 286

25 Query: 181 AEELGREVDADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAI IAKYILEAHKTAGTTL 240  
AEELGR VDADVLVATDPDADRLGVEIRQPDGSY NLSGNQIGAI IAKYILEAHKTAGTTL

Sbjct: 287 AEELGRNVDADVLVATDPDADRLGVEIRQPDGSY NLSGNQIGAI IAKYILEAHKTAGTTL 346

30 Query: 241 PENAALAKSIVSTELVT KIAESYGATMFNVL TGFKFIAEKIQEFEEKHNHTYMF GFEE SF 300  
P NAAL KSIVSTELVT KIAESYGATMFNVL TGFKFI EKI EFE +HN+TYMFGFEESF

Sbjct: 347 PANAALCKSIVSTELVT KIAESYGATMFNVL TGFKFIEGKIHEFEFTQHNYTYMFGFEESF 406

Query: 301 GYLIKPFVRDKDAIQAVLLVAEIAAAYRSRGLT LADGIDEIYKEYGYFAEKTISVTL SGV 360  
GYLIKPFVRDKDAIQAVL+VAEIAAAYRSRG+TLADGI+EIYK+YGYF+EKTISVTL SGV

35 Sbjct: 407 GYLIKPFVRDKDAIQAVLIVAEIAAAYRSRGMTLADGIEEYKQYGYFSEKTISVTL SGV 466

Query: 361 DGAAEIKKIMDKFRENGPKQFNMTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYTLA 420  
DGAAEIKKIMDKFR N PKQFNMTDI EDF +QTAT DG + LTPPSNVLY LA

40 Sbjct: 467 DGAAEIKKIMDKFR RNAPKQFNMTDI AKTEDFLEQTATTADG-VEKLTPPSNVLYILA 525

Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNDLADAETKIANIEKEITTFV 465  
DDSW AVRPSGTEPKIKFYIATVG ADA+ KIANIE EI FV

Sbjct: 526 DDSWFAVRPSGTEPKIKFYIATVG ETEADAKEKIANIEAEINAFV 570

45 There is also homology to SEQ ID 6156:

Query: 1 MAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA 60  
+AQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA

50 Sbjct: 107 LAQHGIKSYVFEALRPTPELSFAVRHLNAYAGIMVTASHNPAPFNGYKVYQDGGQLPPA 166

Query: 61 DADALTD FIRAIENPF FAVELADLDESKSSGLIQVIGEDVDIEYLREV KDVNINQDLINNF 120  
DADALTD FIRAIENPF FAVELADLDE+KSSGLIQVIGEDVD+EYLREV KDVNINQDLINNF

55 Sbjct: 167 DADALTD FIRAIENPF FAVELADLDENKSSGLIQVIGEDVDMEYLREV KDVNINQDLINNF 226

Query: 121 GKDMKIVYTP LHGTGEMLTRRALAQAGFESV VVVSQAKADPDFSTVKSPNPESQA AFAL 180  
GKDMKIVYTP LHGTGEML RRALAQAGFESV VVVSQAKADPDFSTVKSPNPESQA AFAL

Sbjct: 227 GKDMKIVYTP LHGTGEML RRALAQAGFESV VVVSQAKADPDFSTVKSPNPESQA AFAL 286

60 Query: 181 AEELGREVDADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAI IAKYILEAHKTAGTTL 240  
AEELGREV+ADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAI IAKYILEAHKTAGTTL

Sbjct: 287 AEELGREVEADVLVATDPDADRLGVEIRQPDGSYKNLSGNQIGAI IAKYILEAHKTAGTTL 346

Query: 241 PENAALAKSIVSTELVT KIAESYGATMFNVL TGFKFIAEKIQEFEEKHNHTYMF GFEE SF 300  
PENAALAKSIVSTELVT KIAESYGATMFNVL TGFKFIAEKIQEFEEKHNHTYMF GFEE SF

Sbjct: 347 PENAALAKSIVSTELVT KIAESYGATMFNVL TGFKFIAEKIQEFEEKHNHTYMF GFEE SF 406

-2385-

Query: 301 GYLIKPFVRDKDAIQAVLLVAEIAAYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 360  
 GYLIKPFVRDKDAIQAVLLVAEIAAYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV  
 Sbjet: 407 GYLIKPFVRDKDAIQAVLLVAEIAAYRSRGLTLADGIDEIYKEYGYFAEKTISVTLSGV 466

Query: 361 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA 420  
 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA  
 Sbjet: 467 DGAAEIKKIMDKFRENGPKQFNNTDIVLLEDFQKQTATKNDGTISNLTTPPSNVLYKTYLA 526

Query: 421 DDSWIAVRPSGTEPKIKFYIATVGNDLADAETKIANIEKEITTFV 465  
 DDSWIAVRPSGTEPKIKFYIAT+G+ L A+ KIANIE EI TFV  
 Sbjet: 527 DDSWIAVRPSGTEPKIKFYIATIGTDLDAEQKIANIETEINTFV 571

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2119

A DNA sequence (GBSx2235) was identified in *S.agalactiae* <SEQ ID 6541> which encodes the amino acid sequence <SEQ ID 6542>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1564(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9905> which encodes amino acid sequence <SEQ ID 9906> was also identified. There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2120

A DNA sequence (GBSx2236) was identified in *S.agalactiae* <SEQ ID 6543> which encodes the amino acid sequence <SEQ ID 6544>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -9.92 | Transmembrane | 162 - 178 ( 135 - 184) |
| INTEGRAL | Likelihood = -7.11 | Transmembrane | 58 - 74 ( 56 - 78)     |
| INTEGRAL | Likelihood = -6.42 | Transmembrane | 136 - 152 ( 135 - 161) |
| INTEGRAL | Likelihood = -5.20 | Transmembrane | 23 - 39 ( 21 - 49)     |
| INTEGRAL | Likelihood = -1.75 | Transmembrane | 485 - 501 ( 485 - 501) |

----- Final Results -----  
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35376 GB:AE001710 ABC transporter, ATP-binding protein  
 [Thermotoga maritima]  
 Identities = 216/552 (39%), Positives = 336/552 (60%), Gaps = 3/552 (0%)

Query: 26 MALLGTVVQVCLTVYLPVLIGQAVDVVLSPHSMILLPLPIMWKMIIVILANTTIQWINPLL 85  
 M + V L V P LIG+ +DVV P LL M + + +++ W+ +

-2386-

Sbjct: 41 MVFVFTVSSILGVLSPYLIGKTIDVVFVPRRFDLLPRYMLILGTIYALTSLFLWLQGKI 100

Query: 86 YNRLIFHYVASLRKAVMEKLNLLPIAYLDKRGIGDLISRVTTDTEQLSNGLLMVFNQFFV 145  
L V LRK + EKL +P+ + D+ GD+ISRV D + ++N L QFF

5 Sbjct: 101 MLTSLQDVVFRRLRKELFEKLQRPVPGFFDRTPHGDIISRVINDVDNINNVLGNSIIQFFS 160

Query: 146 GLLTIIVTIFSMKIDLLMLFLVLFLTPLSLFLARFIKKS-Y-HLYQNQTASRGRTQFI 204  
G++T+ + M +++++ + L + PL++ + + ++ ++ + Y+NQ G+ I

10 Sbjct: 161 GIVTLGAVIMFRVNVLSLVTLSIVPLTVLITQIVSSQTRKYFYENQVRV-LQLNGII 219

Query: 205 EEMVSQESLIQAFSAQEESDHFTINQYANFSQSAIFYSTVNPSTRFINSLIYGFLLA 264  
EE +S ++I+ F+ +E+ + F +N+ A +S + P +N+L + ++

Sbjct: 220 EEDISGLTVIKLFTREEKEMEKFDRVNESLRKVGTKAQIFSGVLPPLMMVMNVLGFALIS 279

15 Query: 265 GIGALRIMSGAFSVGQLITFLNVNQYTKPFNDISSVLSEMSALACAERLYSILEESSP 324  
G G + +VG + TF+ Y Q+T+P N++S+ + +Q ALA AER++ IL+

Sbjct: 280 GFGGWLALKDIIITVGTIATFIGYSRQFTRPLNELSNQFNMIQMALASAERIFEILDLEE 339

Query: 325 NITGTEKLDSSTVKGQIDFKNVVFGYNKSKLLNGINLHIPAGAKVAIVGPTGAGKSTLI 384  
+ ++ V+G+I+FKNV F Y+K K +L I HI G KVA+VGPTG+GK+T++

20 Sbjct: 340 K-DDPDAVELREVRGEIEFKNVWFSYDKKKPVLKDITPHIKPGQKVALVGPTGSGKTTIV 398

Query: 385 NLIMRFYEVDDGGNILLDCPKITDYEPSQLRQEI GMVLQETWLKSATIHNDNIAYANPKASR 444  
NL+MRFY+VD G IL+D I + S LR IG+VLQ+T L S T+ +N+ Y NP A+

25 Sbjct: 399 NLLMRFYDVRGQILVDGIDIRKIKRSSSIGIVLQDTILFSTTVKENLKYGNPGATD 458

Query: 445 EEVIEAAKAANADFFIKQLPNGYDTYLEDAGDSLQGCQLLTIAIFLKLPRILILDEA 504  
EE+ EAAK ++D FIK LP GY+T L D G+ LSQGC QLL I R FL P+ILILDEA

30 Sbjct: 459 EEIKEAAKLTHSDHFIKHLPEGYETVLTDNGEDLSQGQRQLLAITRAFLANPKILILDEA 518

Query: 505 TSSIDTRTEVLVQEAQMLMKGRSTFIIAHLSTIQTADIILVMVSGEIVEVGNHSELMA 564  
TS++DT+TE +Q A LM+G+TS IIAHL+TI+ AD+I+V+ GEIVE+G H EL+

Sbjct: 519 TSNVDTKTEKSIQAAMWKLMEGKTSIIIAHLNLTIKNADLIIVLRDGEIVEMGKHDELIQ 578

35 Query: 565 QKGIYYQMNAQ 576  
++G YY++ +Q

Sbjct: 579 KRGFYYELFTSQ 590

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6545> which encodes the amino acid  
40 sequence <SEQ ID 6546>. Analysis of this protein sequence reveals the following:

Possible site: 56  
>>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.07 | Transmembrane | 162 - 178 ( 159 - 182) |
| INTEGRAL | Likelihood = -7.17 | Transmembrane | 143 - 159 ( 137 - 161) |
| INTEGRAL | Likelihood = -5.84 | Transmembrane | 23 - 39 ( 19 - 45)     |
| INTEGRAL | Likelihood = -5.68 | Transmembrane | 68 - 84 ( 60 - 86)     |
| INTEGRAL | Likelihood = -2.55 | Transmembrane | 261 - 277 ( 256 - 278) |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.4227(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the databases:

55 >GP:AAD35376 GB:AE001710 ABC transporter, ATP-binding protein  
[Thermotoga maritima]  
Identities = 206/572 (36%), Positives = 342/572 (59%), Gaps = 5/572 (0%)

Query: 2 IKTDHLLKRVLQDLLKKPLPVCILVIASFVQVG--LSVYLPVLIGKAVDMSLSVNSWQT 59  
+K L+R+L L +P +++++ FV V L V P LIGK +D+ +

60 Sbjct: 18 LKNPTATLRRLLGYL--RPHTFTLIMVVFVTVSSILGVLSPYLIGKTIDVVFVPRRFDL 75

Query: 60 LKWLLGQMLVIIIVNTLIQWVMPVLSRLLYQYSQQLKDKLLEKIHLRPFAYLDRQTIGD 119  
L + + I + +L+ W+ + L +L+ +L EK+ R+P + DR GD

65 Sbjct: 76 LPRYMLILGTIYALTSLFLWLQGKIMLTSLQDVVFRRLRKELFEKLQRPVPGFFDRTPHGD 135

-2387-

5 Query: 120 LVSRVITDTEQLINGLQMVFNQFILGLLTILCTIIAMAQIDWMLLILVLVLTSSSLFLAR 179  
 ++SRVI D + + N L QF G++T+ +I M +++ ++ ++ L + P ++ + +  
 Sbjct: 136 IISRVINDVDNINNVLGNSIIQFFSGIVTLGAVIMMFRVNVILSLVTLIVPLTVLITQ 195

10 Query: 180 FIAQKSFHYAQATKSRGNLAQFTEEILRQEGVLQFNQAQESICDYHVLNKTYCEASQK 239  
 ++ ++ Y + G L EE + +++LF +E+ + + +N++ + K  
 Sbjct: 196 IVSSQTRKYFYENQRVLGQLNGIIEEDISGLTVIKLFTREEKEMEKFDRVNESLRKVGTK 255

15 Query: 240 AIFYASTVNPATRFINSVIYALLAGLGAVRIMAGLFSVGQLTTFLNVVVQYTKPFNDISS 299  
 A ++ + P +N++ +AL++G G + + +VG + TF+ Q+T+P N++S+  
 Sbjct: 256 AQIFSGVLPPMLNMVNLGFALISGFGWLALKDIITVGTIATFIGYSRQFTRPLNELSN 315

20 Query: 300 VLAEIQSSLACAQRLYDLLDIEIKEQEHFLTFKASAVKGQIDFEEVSFSYQKDRPLLKDI 359  
 IQ +LA A+R++++LD+E +E++ + V+G+I+F+ V FSY K +P+LKDI  
 Sbjct: 316 QFNMIQMALASAERIFEILDLE-EEKDDPDAVELREVGEIEFKNVWFSYDKKKPVLLKDI 374

25 Query: 360 NFSVPAGSKVAIVGPTGAGKSTLINLLMRFYELDAGSIKLDKVPKCYAKEELRSITGIV 419  
 F + G KVA+VGPTG+GK+T++NLLMRFY++D G I +D + I+ + LRS GIV  
 Sbjct: 375 TFIKPGQKVALVGPTGSGKTTIVNLLMRFYDVRGQILVDGIDIRKIKRSSLSIGIV 434

30 Query: 420 LQETWLKDATVHELIAYGSEESRDEVVAAAKAAHAFFIMQLPKTYDTYLSASDDALSQ 479  
 LQ+T L TV E + YG+ A+ +E+ AAK H+ FI LP+ Y+T L+ + + LSQ  
 Sbjct: 435 LQDTILFSTTVKENLKYGNPGATDEEIKEAAKLTHSDHFIKHLPEGYETVLTNGEDLSQ 494

35 Query: 480 GQLQLLAIARMFLKKPKVLVLDEATSSIDIRTEAVIQEALKELMRGRTSFIIAHLSTIQ 539  
 GQ QLLAI R FL PK+L+LDEATS++D +TE IQ A+ +LM G+TS IIAHRL+TI+  
 Sbjct: 495 GQRQLLAITRAFLANPKILILDEATSNVDTKTEKSIQAAMWKLMEGKTSIIIAHRLNTIK 554

40 Query: 540 SADLILVMDQGRLEVWGTASLMSKNGCYVRL 571  
 +ADLI+V+ G +VE G H L+ K G Y L  
 Sbjct: 555 NADLIIVLRDGEIVEMGKHDELIQKRGFFYEL 586

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 340/566 (60%), Positives = 433/566 (76%)

40 Query: 11 KKLVDLLSKKSLVGMALLGTVVQVCLTVYLPVLIGQAVDVVLSPHSMILLPIMWKMI 70  
 K+++QDLL K V + ++ + VQV L+VYLEPVLIG+AVD+ LS +S L ++ +M+  
 Sbjct: 10 KRVLDQLLKKPLPVCILVIASVFQVGLSVYLPVLIGKAVDMSLSVNSWQTLKWLGLQMLV 69

45 Query: 71 VILANTIIQWINPLLYNRLIFHYVASLRKAVMEKLNLLPIAYLDRKIGDGLISRVITDTE 130  
 +I+ NT+IQW+ PL+Y+RL++ Y L+ ++EK++ LP AYLD++ IGD+SRV TDTE  
 Sbjct: 70 IIVVNTLIQWVMPVLSRLLYQYSQQLKDKLLEKIHRLPFAYLDRQTIGDLVSRVITDTE 129

50 Query: 131 QLSNGLLMVFNQFFVGLLTIIIVTIFSMKIDLLMLFLVLFLTPLSLFLARFIAKKS+YHLY 190  
 QL NGL MVFNQF +GLLTI+ TI +MA+ID LML LVL LTP SLFLARFIA+KS+H  
 Sbjct: 130 QLINGLQMVFNQFILGLLTILCTIIAMAQIDWMLLILVLVLTSSSLFLARFIAQKSFHYA 189

55 Query: 191 QNQTASRGRQTQFIEEMVSQESLIQAQSAQEESDHFRITINQYANFSQSAIFYSTVNP 250  
 Q QT SRG QF EE++ QE L+Q F+AQE+S + +N+ Y SQ AIFY+STVNP  
 Sbjct: 190 QAQTKSRGNLAQFTEEILRQEGVLQFNQAQESICDYHVLNKTYCEASQKAIIFYASTVNP 249

60 Query: 251 STRFINSILIYGFAGLAGALRIMSGAFSVGQLITFLNYVNQYTKPFNDISSVLSEMQSALA 310  
 +TRFINS+IY LAG+GA+RIM+G FSVGQL TFLN V QYTKPFNDISSVL+E+QS+LA  
 Sbjct: 250 ATRFINSVIYALLAGLGAVRIMAGLFSVGQLTTFLNVVVQYTKPFNDISSVLAEIQSSLA 309

65 Query: 311 CAERLYSILEESSPNITGTTEKLDSSVTKGQIDFKNVVFYGNKSKLLNLNGLHIPAGAKV 370  
 CA+RLY +L+ +S VKGQIDF+ V F Y K + LL IN +PAG+KV  
 Sbjct: 310 CAQRLYDLLDIEIKEQEHFLTFKASAVKGQIDFEEVSFSYQKDRPLLKDNFSVPAGSKV 369

Query: 371 AIVGPTGAGKSTLINLIMRFYEVDDGNILLDCKPITDYEPSQLRQEIGMVLQETWLKAT 430  
 AIVGPTGAGKSTLINL+MRFYE+D G+I LD PI Y +LR G+VLQETWLK AT  
 Sbjct: 370 AIVGPTGAGKSTLINLIMRFYELDAGSIKLDKVPKCYAKEELRSITGIVLQETWLKAT 429

Query: 431 IHDNIAIYANPKASREEVIEAAKAANADFFIKQLPNGYDTYLEDAGDSLSQGCQLLTAR 490  
 +H+ IAY + +ASR+EV+ AAKAA+A FFI QLP YDTYL + D+LSQGC QLL IAR  
 Sbjct: 430 VHELIAYGSEESRDEVVAAAKAAHAFFIMQLPKTYDTYLSASDDALSQGCQLLAIAR 489

-2388-

Query: 491 IFLKLPRIILDEATSSIDTRTEVLVQEAQMLMKGRTSFIIAHLSTIQTADIILVMVS 550  
 +FLK P++L+LDEATSSID RTE ++QEA + LM+GRTSFIIAHLSTIQ+AD+ILVM  
 Sbjct: 490 MFLKKPKVLVLDEATSSIDIRTEAVIQEALKELMRGRTSFIIAHLSTIQSADLILVMDQ 549

Query: 551 GEIVEVGNHSELMAQKGIYYQMQNAQ 576  
 G +VE G H+ LM++ G Y ++Q +  
 Sbjct: 550 GRLVEWGTHASLMSKNGCYVRLQKIE 575

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2121

A DNA sequence (GBSx2237) was identified in *S.agalactiae* <SEQ ID 6547> which encodes the amino acid sequence <SEQ ID 6548>. Analysis of this protein sequence reveals the following:

15 Possible site: 26  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1099 (Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2122

A DNA sequence (GBSx2238) was identified in *S.agalactiae* <SEQ ID 6549> which encodes the amino acid sequence <SEQ ID 6550>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

35 INTEGRAL Likelihood = -13.69 Transmembrane 157 - 173 ( 130 - 182)  
 INTEGRAL Likelihood = -10.88 Transmembrane 56 - 72 ( 49 - 77)  
 INTEGRAL Likelihood = -7.75 Transmembrane 239 - 255 ( 235 - 258)  
 INTEGRAL Likelihood = -6.42 Transmembrane 133 - 149 ( 130 - 156)  
 INTEGRAL Likelihood = -4.78 Transmembrane 271 - 287 ( 270 - 289)  
 INTEGRAL Likelihood = -1.91 Transmembrane 20 - 36 ( 20 - 37)

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6477 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 45 The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD35375 GB:AE001710 ABC transporter, ATP-binding protein  
 [Thermotoga maritima]  
 Identities = 196/570 (34%), Positives = 327/570 (56%), Gaps = 5/570 (0%)

50 Query: 1 MKRLTYFFKGYIKETIFGPLFKLLEASFELLVPIVIAKMIDETIPRGDRSGLLLQIGLIF 60  
 MK L Y K Y + PLF ++E +L P ++A+++DE I RGD S L+L+ G++  
 Sbjct: 1 MKTLARYLKPYWIFAVLAPLFMVVEVICDLSQPTLLARIVDEGIARGDFS-LVLKTGILM 59

Query: 61 FLAA-VGVVVAITAQYYSSKAAVGYTRQLTEDLYQKVM SLGKKDRDELGTASLITRLTAD 119

-2389-

+ A +G V I ++S A+ + L DL++KV+S + + T+SLITRLT D  
 5 Sbjet: 60 LIVALIGAVGGIGCTVFASYASQNFADLRDLFRKVLSPSISNVNRFHTSSLITRLTND 119  
 Query: 120 TFQIQTGLNQFLRLFLRAPIIVFGAIIAFSISPSLTIWFLVMVVTLFIIVFVMSRLLNP 179  
 Q+Q + LR+ +RAP++ G I+MA SI+ L+ + ++ + ++ +++ NP  
 Sbjet: 120 VTQLQNLVMMLLRIVVRAPLLFVGGIVMAVSINVKLSSVLIFLIPPIVLLFVWLTKKGNP 179  
 Query: 180 IYLKIRTSTDYLVKLTRQQLQGVVRIRAFNQVDRESEAFNDINYHTNLQKAGRLSSLV 239  
 ++ KI+ STD + ++ R+ L GVRV+RAF + + E+E F N + A L  
 10 Sbjet: 180 LFRKIQESTDEVNRVRENLLGVRVVRAFRREEYENENFRKANESLRRSIISAFSLIVFA 239  
 Query: 240 TPLTFLVVNITLVVVIWRGNLNIANHLLSQGMLVALINYLQILVELLKMMLVTSLNQS 299  
 PL +VN+ ++ ++W G + + N+ + G ++A NYL+QI+ L+ + ++ + ++  
 15 Sbjet: 240 LPLFIFIVNMGMIAVLWFGGVLVRNNQMEIGSIMAYTNYLMQIMFSLMMIGNILNFIVRA 299  
 Query: 300 YISAKRIIAVF-ERPS-EIIDDKLEPKYSNKALEVQEMAFSYPNSSEKALSDITFSMNVG 357  
 SAKR++ V E+P+ E D+ L ++ + + F Y +++ LS + FS+ G  
 Sbjet: 300 SASAKRVLEVLNEKPAIEEADNALALEPNEGVSFENVEFRYFENTDPVLSGVNFSVKPG 359  
 Query: 358 ETLGIIGGTGSGKSTLINLLHHYKQVEGDIDYHQKSPDTISNWRTLVRVVPQNAQLF 417  
 + ++G TSGKSTL+NL+ + + G +++ + + R + VPQ LF  
 20 Sbjet: 360 SLVAVLGETGSGKSTLMNLIPRLIDPERGRVEVDELVDVRTVKLKDRLRGHISAVPQETVLF 419  
 Query: 418 KGTIRSNLSLGLGVSEEKLTALAEIAQASDFVKEKDGQLDAPVESFGRNFSGGQRQLT 477  
 GTI+ NL G +++++ A +IAQ DF+ D+ VE GRNFSGGQ+QRL+  
 25 Sbjet: 420 SGTIKENLKWGREDATDDEIVEAAKIAQIHDFIISLPEGYDSRVERGGRNFSGGQKQRLS 479  
 Query: 478 IARALVQDKIPFLILDATSALDYLTEARLFKAITKHFNQTNLIIVSQRINSIQNADRIL 537  
 IARALV+ K LILDD TS++D +TE R+ + ++ I++Q+I + AD+IL  
 30 Sbjet: 480 IARALVK-KPKVLILDDCTSSVDPITEKRILDGLKRYTKGCTTFIITQKIPTALLADKIL 538  
 Query: 538 LLDKKGKQVGFNDHQSLLAHNKVYKSIYHSQ 567  
 +L +GK GF H+ LL H K Y+ IY SQ  
 35 Sbjet: 539 VLHEGKVAGFGTHKELLEHCCKPYREIYESQ 568

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6551> which encodes the amino acid sequence <SEQ ID 6552>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence  
 40 INTEGRAL Likelihood = -12.47 Transmembrane 157 - 173 ( 149 - 185)  
 INTEGRAL Likelihood = -7.75 Transmembrane 55 - 71 ( 51 - 74)  
 INTEGRAL Likelihood = -4.25 Transmembrane 239 - 255 ( 237 - 260)  
 INTEGRAL Likelihood = -3.77 Transmembrane 20 - 36 ( 19 - 37)  
 45 INTEGRAL Likelihood = -3.50 Transmembrane 271 - 287 ( 270 - 288)  
 INTEGRAL Likelihood = -2.55 Transmembrane 133 - 149 ( 130 - 151)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5989(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

!GB:AL137187 putative ABC transporter [Streptomyces ... 296 6e-79  
 55 >GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces  
 coelicolor A3(2)]  
 Identities = 185/569 (32%), Positives = 306/569 (53%), Gaps = 8/569 (1%)  
 Query: 1 MKRLRPYVKGYLKESILGPLFKLLEALFELLVPLLIANMIDISISQHNSQGILRVVLTFLF 60  
 ++ LR Y++ Y K L + L+ L +P L A++ID + + +S IL +  
 60 Sbjet: 3 IRLRLTYLRPYKKPIALLVALQFLQTCASLYLPTLNAHIIDEGVVKGDSGYILSYGALMI 62  
 Query: 61 GLATIGLLLSVTAQYFSSKAAVGFTRQMTDDLFFKKIMFLSKEDQDHLGYASLLSRLTSDS 120  
 G++ ++ ++ A ++ ++ A R + +F ++ S + H G SL++R T+D  
 65 Sbjet: 63 GISLAQVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDV 122

-2390-

5 Query: 121 FQIQGTGINQFLRLFLRAPIIVCGAMVMAYWISPSLTLWFMVMMVIVLLTLVFMVSHLLGPL 180  
 Q+Q L + API+ G +VMA + L+ + +V VL V ++ L PL  
 Sbjct: 123 QQVQMLALMTFTLMVSAPIMCVGGIVMALGLDVPLSGVLLGVVPLAICVTLIVRKLRLPL 182

10 Query: 181 YLLIRRETDHLVRLTSQQLOGIRVIKAFNQTKELQAFKQONMLLSRHQYQAATLANVLN 240  
 + ++ D + R+ +Q+ G RVI+AF + + E Q F++ N L+ L ++  
 Sbjct: 183 FRKMQVRLDTVNRVLRREQITGNRVIRAFVRDEYEQQRFKANTELTEVALGTGNLLALMF 242

15 Query: 241 PMTFLVNLTLILLIWIQGSQVAHRSLSQGMVALINYLQILAECLKMTMLMGTINQSV 300  
 P+ VVNL+ + ++W G+ ++ + G L A + YL+QI+ ++ T + + ++  
 Sbjct: 243 PVMVTVNLSIAVWVGAHRIDSGGMQIGDLTAFLAYLMQIVMSVMMATFMFMVPRAE 302

20 Query: 301 TAAKRINQVFVLADEAPLPLLDKGPISHT-LLTIRHLTFTYPGAEPSTLYDIQLSADQGE 359  
 A+RI +V P+ + H L IR F YPGA EP L I L A GE  
 Sbjct: 303 VCAERIQEVLETESSVPPVAPVTELRHGHLEIREAGFRYPGAEEPVLRHIDLVARPGE 362

25 Query: 360 WIGIIGGTGAGKTTLLIDLICQTSYSQSGEISLW---QGEVPKLTTEWRNVIALVLPQKAQ 416  
 +IG TG+GK+TL+ L+ + + GE+ +N + PKTL + V++LVPQK  
 Sbjct: 363 TTAVIGSTGSGKSTLLGLVPRLEFADTDEVLVNGVDVRTVDPKTLAK--VVSLVPQKPY 419

30 Query: 417 LFKGTIRSNLLLGQSMPISEELWRALELAQAKEFVAALPEQLEAPVEAFGRHFSGGQRQ 476  
 LF GT+ +NL G + +DEELW AL +AQAKEFV+ L L+AP+ G + SGGQRQ  
 Sbjct: 420 LFAGTVATNLRYG-NPDATDEELWHALEVAQAKEFVSELEGGDLAPIAQGGTNVSGGQRQ 478

35 Query: 477 RLAIARALLKPKPILILDDASSALDNETRGRLEFKALKEELSDVLVILVTQSIKNLQFADK 536  
 RLAIAR L++ I + DD+ SALD T L L +E ++ V++V Q + ++ AD+  
 Sbjct: 479 RLAIARTLVQRPEIYLFDDSFALDYATDAALRAELAQETAETVIVIVAQRVATIRDADR 538

40 Query: 537 ILVLEQGHQLDFASHDQLKVSNALYQEML 565  
 I+VL++G + H +L N Y+E++  
 Sbjct: 539 IVVLDEGRVVGVRHHELMADNETYREIV 567

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 313/568 (55%), Positives = 428/568 (75%), Gaps = 9/568 (1%)

40 Query: 1 MKRLTYYPKGYIKETIFGFLFKLEASFELLVPIVIAKMIDETIPGRDRSGLLLQIGLIF 60  
 MKRL Y KGY+KE+I GPLFKLEA FELLVP++IA MID +I + + G+L + +F  
 Sbjct: 1 MKRLRPYVKGYLKESILGPLFKLEALFELLVPLLIANMIDISISQHSQGILRVVLTFL 60

45 Query: 61 FLAAGVVVAITAQYYSKAAVGYTRQLTEDLYQKVMISLGGKDRDELGTASLITRLTADT 120  
 LA +G++++TAQY+SSKAAVG+TRQ+T+DL++K+M L K+D+D LG ASL++RLT+D+  
 Sbjct: 61 GLATIGLLLSVTAQYFSSKAAVGFTQMTDDLFFKIMFLSKEDQDHLGYASLLSRLTSDS 120

50 Query: 121 FQIQGTGLNQFLRLFLRAPIIVFGAIIAFAISPSLTIWFLVMVVTFLFIIVFMSRLLNPI 180  
 FQIQGTG+NQFLRLFLRAPIIV GA++MA+ ISPSLT+WF++MV+ L +VFMV LL P+  
 Sbjct: 121 FQIQGTGINQFLRLFLRAPIIVCGAMVMAYWISPSLTLWFMVMMVIVLLTLVFMVSHLLGPL 180

55 Query: 181 YLKIRTSTDYLVKLTRQQLQGVIRAFNQVDRESEAFNDINYHYTNLQLKAGRLSSSLVT 240  
 YL IR TD+LV+LT QQLQG+RVI+AFNQ +E +AF N + Q +A L++++  
 Sbjct: 181 YLLIRRETDHLVRLTSQQLOGIRVIKAFNQTKELQAFKQONMLLSRHQYQAATLANVLN 240

60 Query: 241 PLTFLVNNITLVVLIWRGNLIANHLLSQGMVALINYLQILVELLKMMLVTSNLQSY 300  
 P+TFLVNN+TL+++IW+G+ +A+ LSQGMVALINYLQIL ELLKMML+ ++NQS  
 Sbjct: 241 PMTFLVNLTLILLIWIQGSQVAHRSLSQGMVALINYLQILAECLKMTMLMGTINQSV 300

65 Query: 301 ISAKRIIAVF---ERPSEIIDDLEPKYSNKALEVQEMAFSYPNSSEKALS DITFSMN 356  
 +AKRI VF E P ++ D S L ++ + F+YP ++E +L DI S +  
 Sbjct: 301 TAAKRINQVFVLADEAPLPLLDK---GPISHTLLTIRHLTFTYPGAEPSTLYDIQLSADQ 357

70 Query: 357 GETLGIIGGTGSGKSTLINLLHIIKYQEGDIDIYHQKSPDTISNWRTLVRVVPQNAQL 416  
 GE +GIIGGTG+GK+TLI+L+ Y G+I + QG+ P T++ WR ++ +VPQ AQL  
 Sbjct: 358 GEWIGIIGGTGAGKTTLLIDLICQTSYSQSGEISLWQGEVPKLTTEWRNVIALVLPQKAQL 417

75 Query: 417 FKGITIRSNLSGLG-KVSEKLTALAEIAQASDFVKEKQGLDAPVESFGRNFSGGQRQR 475  
 FKGITIRSNL LG +S+E+LW ALE+AQA +FV QL+APVE+FGR+FSGGQRQR  
 Sbjct: 418 FKGITIRSNLLGQSMPISEELWRALELAQAKEFVAALPEQLEAPVEAFGRHFSGGQRQR 477



-2391-

Query: 476 LTIARALVQDKIPFLILDDATSALDYLTPEARLFKAITKHFNQTNLIIVSQRINSIQNADR 535  
 L IARAL++ K P LILDDA+SALD T RLFKA+ + + +I+V+Q I ++Q AD+  
 Sbjct: 478 LAIARALLKPK-PILILDDASSALDNETRGRLEFKALKEELSDVLVILVTQSIKNLQFADK 536

Query: 536 ILLLDKKGKQVGFNDHQSLLAHNKVYKSI 563  
 IL+L++G Q+ F +H L N +Y+ +  
 Sbjct: 537 ILVLEQGHQLDFASHDQLKVSNALYQEM 564

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2123

A DNA sequence (GBSx2239) was identified in *S.agalactiae* <SEQ ID 6553> which encodes the amino acid sequence <SEQ ID 6554>. Analysis of this protein sequence reveals the following:

15 Possible site: 43  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -12.26 Transmembrane 8 - 24 ( 1 - 28)

20 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAB84433 GB:AF027868 RAS-related protein [Bacillus subtilis]  
 Identities = 53/140 (37%), Positives = 78/140 (54%), Gaps = 2/140 (1%)

Query: 28 VKKVLQYHDLVQNTLAENGSEANVHLVLSMIYTETKGAIDVMQSSESISGTTNSITDSH 87  
 ++++ Y LV+ L G L+L M+Y E+KG D MQSSES+ N ITD  
 30 Sbjct: 49 LERLTDYKPLVEEELESQGLSNYTSLLILGMMYQESKKGNDPMQSSESLGLKRNEITDPQ 108

Query: 88 TSIKHGVTLLSQNISQAKKAKVDVWTVAVQAYNFGSSYIDYVADHGGENSIELAKNYSKNV 147  
 S+K G+ + K+ VD+ T +Q+YN G+ YID+VA+HGG ++ ELAK YS+  
 35 Sbjct: 109 LSVKQGIKQFTLMYKTGKEKGVLDLTIIQSYNMGAGYIDFVAEHGGTHTTEELAKQYSEQQ 168

Query: 148 VA--PSLGNNGDITYFYHP 165  
 V P L G+ + +P  
 Sbjct: 169 VKKNPDLYTCGGNAKNFRYP 188

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4143> which encodes the amino acid sequence <SEQ ID 4144>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have an uncleavable N-term signal seq  
 45 INTEGRAL Likelihood = -2.66 Transmembrane 8 - 24 ( 7 - 25)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2062(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 134/200 (67%), Positives = 165/200 (82%), Gaps = 1/200 (0%)

55 Query: 1 MFKFLKRLIALIIIIIFIGYRLVIIHENVKKVLQYHDLVQNTLAENGSEANVHLVLSMIYT 60  
 MF+ LKR + +++ F+ Y+ +IH NV++VL Y +V+ TLAEN ++ANV LVL+MIYT  
 Sbjct: 1 MFRLLKRACSFLLL-FVIYQSFVIHNVQVRVLA YKPMVEKTLAENDTKANVDLVLAMIYT 59

Query: 61 ETKGDAIDVMQSSESISGTTNSITDSHTSIKHGVTLLSQNISQAKKAKVDVWTVAVQAYNF 120  
 ETKG DVMQSSES SG NSITDS SI+HGV LLS N++ A++A VD WTVAVQAYNF

-2392-

Sbjct: 60 ETKGGEADVMSSESSSGQKNSITDSQASIEHGVNLLSHNLALAEAGVDSWTAVQAYNF 119

Query: 121 GSSYIDYVADHGGENSIELAKNYSKNVAPSLGNYNGDTYFYHPLALISGGKLYKNGGN 180  
 G++YIDY+A+HGG+N+++LA YSK VVAPSLGN +G TYFYHPLALISGGKLYKNGGN

5 Sbjct: 120 GTAYIDYIAEHGGQNTVDLATTYSKTVVAPSLGNTSGQTYFYHPLALISGGKLYKNGGN 179

Query: 181 IYYSREVQFNLYLIKIMELF 200  
 IYYSREV FNLYLI++M LF

10 Sbjct: 180 IYYSREVFHFNLYLIELMSLF 199

SEQ ID 6554 (GBS244) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 4; MW 23.1kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 2; MW 48kDa).

GBS244-GST was purified as shown in Figure 211, lane 5.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2124

A DNA sequence (GBSx2240) was identified in *S.agalactiae* <SEQ ID 6555> which encodes the amino acid sequence <SEQ ID 6556>. Analysis of this protein sequence reveals the following:

20 Possible site: 38  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2401(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9837> which encodes amino acid sequence <SEQ ID 9838> was also identified.

- 30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB71302 GB:AJ130879 hypothetical protein [Clostridium  
 sticklandii]  
 Identities = 32/95 (33%), Positives = 53/95 (55%), Gaps = 1/95 (1%)

35 Query: 235 LSPEKLADQLFDDNLTARLTFVDELKDAIPGPVQVSDIDHSRQIKKLENQKLSLSNGTEL 294  
 LS EK + F++ + + + L A Q+ ++ + +K E QK+ +GIE+

Sbjct: 2 LSVEKALETAFEETDEIKAIYKEALSKAGIENEQI-EVSETALKRKFEIQKIITESGIEV 60

Query: 295 IVPNNVYQDAESVEFIQNPDGTSILIKNIQDIQN 329  
 +P N Y D +EF+ N DGT S++IKNI +IQ+

40 Sbjct: 61 KIPVNYGDPKLEFVANGDGTVSLVIKNIGNIQS 95

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6557> which encodes the amino acid sequence <SEQ ID 6558>. Analysis of this protein sequence reveals the following:

45 Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.3336(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2393-

Identities = 246/325 (75%), Positives = 286/325 (87%)

Query: 6 MMDFYIKQIIHQFSPNDTELVLSDTPLTLTPRIDYFRKKLSKVFSDEAKRGYFGEDNV 65  
 M+D YIK+I+IHQFSPNDTEL+LSD +++TPRID+YFRKKL+KVFSDEAKRG F +N  
 5 Sbjct: 1 MLDSYIKRIVIHQFSPNDELLSDRLVSITPRIDEYFRKKLAKVFSDEAKRGQFEANNT 60

Query: 66 FMSHLQDDLYVSSCQIAQLWKEEFVISEDQKTNDLVFIQFDKDGMEHFAFLRISLKEQFA 125  
 F + + DDL +S IAQLWKE FVISEDQKTNDLVF+QFDKDG FAFLRI+LKEQFA  
 10 Sbjct: 61 FFTTIGDDLLTSTVTIAQLWKEAFVISEDQKTNDLVFVQFDKDGEPFFAFLRIALKEQFA 120

Query: 126 HVSENQEQPITITQNNLPSAAQTPDEALVVKSSKQYYLIEKRIKHNGSFANYFSENLLQ 185  
 H+S+N E P T+TQNNLPS QTPDEALV+N S QYYLIEKR+KHNGSFANYFSE+LL+  
 Sbjct: 121 HLSDNYEHPFTVTQNNLPSPTQTPDEALVINLKSQYYLIEKRVKHNGSFANYFSEHLLK 180

Query: 186 VQPEQSVKKSIIKMVEQTAQKIAENFNKDDFSQSKMKSATYKNLEEEQELSPEKLADQLF 245  
 V PEQSVKKSIIKM+EQTAQKIAE+FN+DDF+FQSKMKS ++K LE + LSPEKLADQLF  
 15 Sbjct: 181 VTPEQSVKKSIIKMIEQTAQKIAEHFNQDDFTFQSKMKS+LFLKQLEADDVLSPKADQLF 240

Query: 246 DDNLARLTFVDELKDAIPGPVQVSDIDHSRQIKKLENQKLSLSNGIELIVPNNVYQDAE 305  
 DDNLARLTFVD++KD IP P+++SDI+HSRQIKKLENQKLSLSNGIEL VPN +YQDAE  
 20 Sbjct: 241 DDNLARLTFVDQVKDVIPEPIKISDIEHSRQIKKLENQKLSLSNGIELTVPNATYQDAE 300

Query: 306 SVEFIQNPDTYSILIKNIQDIQNK 330  
 +VEF+ N DGTYSILIKNI+DI+ K  
 25 Sbjct: 301 AVEFLNDDGTYSILIKNIEDIKTK 325

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2125

30 A DNA sequence (GBSx2241) was identified in *S.agalactiae* <SEQ ID 6559> which encodes the amino acid sequence <SEQ ID 6560>. This protein is predicted to be Serine hydroxymethyltransferase (glyA-1). Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3876(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35802 GB:AE001743 serine hydroxymethyltransferase [Thermotoga maritima]  
 Identities = 243/416 (58%), Positives = 307/416 (73%), Gaps = 7/416 (1%)

45 Query: 9 KEFDQELWQAIHDEEIRQNNIELIASENVVSKAVMAAQGSVLTKNYAEGYPSHRYGGT 68  
 K+ D E+++ + +E RQ+ +ELIASEN S AV+ GS+LTKNYAEGYP RYGG  
 Sbjct: 6 KQVDPEIYEVLVNELKRQEYGLELIASENFASLAVIETMGSMLETKNYAEGYPKKRYGGC 65

50 Query: 69 DCVDVVESLAIERAKTLFNAEFANVQPHSGSQANAAAYMALIEPGDTVLGMDLAAGGHLT 128  
 + VD E AIERAK LF A+FANVQPHSGSQAN A Y+AL +PGDT++GM L+ GGHLT  
 Sbjct: 66 EWVDRAEERAIERAKRLFGAKFANVQPHSGSQANMAVYLALAQPGDTIMGMSLSHGGLT 125

55 Query: 129 HGASVSFSGKTYHFVSYSVDPKTEMLDYDNILKIAQETQPKLIVAGASAYSRIIDFEKFR 188  
 HGA V+PSGK + V Y V+ +TE +DYD + ++A E +PK+IVAG SAY+RIIDF++FR  
 Sbjct: 126 HGAPVNFSGKIFKVVPYGVNLETETIDYDEVRRLLALEHKPKIIVAGGSAYARIIDFKRFR 185

60 Query: 189 QIADAVDAYLMVMDMAHIAGLVASGHHPSPIYAHVTTTTTHKTLRGPRGGLILTNDIA 248  
 +IAD V AYLMVMDMAH AGLVA+G HP+P+ YAHV T+THKTLRGPRGGLILTND IA  
 Sbjct: 186 ETADVGAYLMVMDMAHFAGLVAAGIHPNPLEYAHVVTSTTHKTLRGPRGGLILTNDPEIA 245

Query: 249 KKINSVFPGLQGGPLEHVIAAKAVALKEALDPSFKIYGEDIKNAQAMAKVFKEDDDDFH 308

-2394-

K ++ +FPG+QGGPL HVIAAKAV KEA+ FK Y + ++KNA+ MA+ F++ +  
 Sbjct: 246 KAVDKTIFPGIQGGPLMHVIAAKAVCFKEAMTEEFKEYQKQVVKNAKMAEEFQK-RGYR 304

Query: 309 LIISDGTDNHFLVDVTKVIENGKKAQNVLEEVNITLNKNSIPFERLSPFKTSGIRIGTPA 368  
 ++S GTD HLFLVD+T GK A+ LE IT+NKN+IP E+ SPF SGIRIGTPA  
 Sbjct: 305 IVSGGTDTHLFLVDLTPKIDITGKAAEKALESCGITVNKNTIPNEKRSPFVASGIRIGTPA 364

Query: 369 ITSRCMGVEESRRIAELMIKALKN--HENQDVLTEVRQE----IKSLTDAFPLYEN 418  
 +T+RGM EE IAE++ L N EN V EVR+E ++ L + FPLY +  
 Sbjct: 365 VTTRGMKEEEMEEIAEMIDLVLNSVIDENGTVKPEVREEVSKKVRCLCERFPLYRD 420

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6561> which encodes the amino acid sequence <SEQ ID 6562>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.00 Transmembrane 196 - 212 ( 196 - 212)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15707 GB:Z99122 serine hydroxymethyltransferase [Bacillus subtilis]  
 Identities = 250/407 (61%), Positives = 311/407 (75%), Gaps = 2/407 (0%)

Query: 14 DKELWDIAHAEERQEHHELIASENMVSKAVMAAQGSVLTKNYAEGYPGNRYGGTECV 73  
 D+++++AI E ERQ+ IELIASEN VS+AVM AQGSVLTKNYAEGYPG RYYGG E V  
 Sbjct: 8 DEQVFNAIKNERERQOTKIELIASENFVSEAVMEAQGSVLTKNYAEGYPGKRYGGCEHV 67

Query: 74 DIVETLAIERAKKLFGAAPFANVQAHSQSQANAAAYMALIEAGDVTVLGMDLAAGGHLTHGS 133  
 D+VE +A +RAK++FGA NVQ HSG+QAN A Y ++E GDTVLMG+L+ GGHLTHGS  
 Sbjct: 68 DVVEDIARDRAKEIFGAHVNVQPHSGAQANMAVYFTILEQQGDTVLMGNLSHGGHLTHGS 127

Query: 134 PVNFSGKTYHFVGYSDTDTTEMLNIEAILEQAKAVQPKLIVAGASAYSRSIDFEKFRAIA 193  
 PVNFSG Y+VY Y VD +T+ ++Y+ + E+A A +PKLIVAGASAY R+IDF+KFR IA  
 Sbjct: 128 PVNFSGVQYNFVEYGVDKETQYIDYDDVREKALAHKPKLIVAGASAYPRTIDFKKFREIA 187

Query: 194 DHVGAYLMVMDMAHIAAGLVAAAGVHPSVPVYAHIVTSTTHKTLRGPRGGLILTNDALAKKI 253  
 D VGAY MVDMAHIAAGLVAAAG+HP+PVPYA VT+TTHKTLRGPRGG+IL +E KKI  
 Sbjct: 188 DEVGAYFMVMDMAHIAAGLVAAAGLHPNPVPYADFVTTTTHKTLRGPRGMILCREE-FGKKI 246

Query: 254 NSAVFPGLQGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMAAVFAQDDRFRLIS 313  
 + ++FPG+QGGPL HVIAAKAV+F E L FK YAQ +I N +A ++ +L+S  
 Sbjct: 247 DKSIFPGIQGGPLMHVIAAKAVSFGEVLQDDFKTYAQNVISNAKRLAEALTKEG-IQLVS 305

Query: 314 GGTDNHFLVDVTKVIANGKLAQNLLDEVNITLNKNAIPFETLSPFKTSGIRIGCAITS 373  
 GGTDNH+ LVD+ + GK+A+++LDE+ IT NKNAIP++ PF TSGIR+G AA+TS  
 Sbjct: 306 GGTDNHLILVDLRLSLGLTGKVAEHVLDLDEIGITSNKNAIPYDPEKPFVTSGIRLGTAAVTS 365

Query: 374 RGMGVKESQTIARLIKALVNHQDETILEEVRQEVRLTDAFPLYKK 420  
 RG + + +I AL NH+ E LEE RQ V LTD FPLYK+  
 Sbjct: 366 RGFDGDALEEVGAILALALKNHDEGKLEEARQVAALTDKFFPLYKE 412

An alignment of the GAS and GBS proteins is shown below.

Identities = 330/417 (79%), Positives = 368/417 (88%)

Query: 1 MIFDKDNKFEDQELWQAIHDEEIRQNNIELIASENVVSKAVMAAQGSVLTKNYAEGYP 60  
 MIFDK N ++FD+ELW AIH EE RQ+++IELIASEN+VSKAVMAAQGSVLTKNYAEGYP  
 Sbjct: 3 MIFDKGNVEDFDKELWDIAHAEERQEHHELIASENMVSKAVMAAQGSVLTKNYAEGYP 62

Query: 61 SHRYGGTDCVDVVESLAIERAKTLFNAEFANVQPHSGSQANAAAYMALIEPGDVTLMGM 120  
 +RYYGGT+CVD+VE+LAIERAK LF A FANVQ HSGSQANAAAYMALIE GDTVLMGM

-2395-

5  
10  
15  
20

Sbjct: 63 GNRYYGGTECVDIVETLAIERAKKLFGAAFANVQAHSGSQANAAAYMALIEAGDTVLGMD 122

Query: 121 LAAGGHLTHGASVSFSGKTYHFVSYSVDPKTEMLDYDNILKIAQETQPKLIVAGASAYS 180  
LAAGGHLTHG+ V+FSGKTYHFV YSVD TEMPL+Y+ IL+ A+ QPKLIVAGASAYS

Sbjct: 123 LAAGGHLTHGSPVNFSGKTYHFVGYSVDTDTTEMLNYEAILLEQAKAVQPKLIVAGASAYS 182

Query: 181 IIDFEKFRQIADAVDAYLMVMDMAHIAGLVASGHHPSPPIYAHVTTTTTHKTLRGPRGGLI 240  
IDFEKFR IAD V AYLMDMAHIAGLVA+G HPSP+PYAH+ T+TTHKTLRGPRGGLI

Sbjct: 183 SIDFEKFRAIADHVGAYLMVMDMAHIAGLVAAGVHPSPVPYAHIVTSTTHKTLRGPRGGLI 242

Query: 241 LTNDDEAIAKKINSVFPGLQGGPLEHVIAAKAVALKEALDPSFKIYGEDIKNAQAMAKV 300  
LTNDDEA+AKKINSVFPGLQGGPLEHVIAAKAVA KEALDP+FK Y + II N AMA V

Sbjct: 243 LTNDDEALAKKINSVFPGLQGGPLEHVIAAKAVAFKEALDPAFKDYAQAIIDNTAAMAAV 302

Query: 301 FKEDDDFHLISDGTNDNHLFLVDVTKVIENGKKAQNVLEEVDNITLNKNSIPFERLSPFKTS 360  
F +DD F LIS GTDNH+FLVDVTKVI NGK AQN+L+EVNITLNKN+IPFE LSPFKTS

Sbjct: 303 FAQDDRFRLLISGGTDNHFVFLVDVTKVIANGKLAQNLLDEVNITLNKNAIPFETLSPFKTS 362

Query: 361 GIRIGTPAITSRGMGVEESRRIELMIKALKNHENQDVLTEVRQEIKSLTDAFPLYE 417  
GIRIG AITSRGMGV+ES+ IA L+IKAL NH+ + +L EVRQE++ LTDAFPLY+

Sbjct: 363 GIRIGCAAITSRGMGVKESQTIARLIKALVNHDQETILEEVRQEVRQLTDAFPLYK 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 2126

A DNA sequence (GBSx2242) was identified in *S.agalactiae* <SEQ ID 6563> which encodes the amino acid sequence <SEQ ID 6564>. Analysis of this protein sequence reveals the following:

Possible site: 30  
>>> Seems to have no N-terminal signal sequence

30

----- Final Results -----

|                         |                               |         |
|-------------------------|-------------------------------|---------|
| bacterial cytoplasm --- | Certainty=0.2289(Affirmative) | < succ> |
| bacterial membrane ---  | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial outside ---   | Certainty=0.0000(Not Clear)   | < succ> |

35

A related GBS nucleic acid sequence <SEQ ID 9839> which encodes amino acid sequence <SEQ ID 9840> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40

>GP:AAD35934 GB:AE001752 conserved hypothetical protein [Thermotoga maritima]  
Identities = 71/198 (35%), Positives = 114/198 (56%), Gaps = 4/198 (2%)

Query: 1 MNDLGQILEDHGAVIMPTETVYGI FAKALSEEAVNHVYELKKRPRDKAMNLNICDFETIL 60  
+ + ++L + +I PTETVYGI A A +EEA +++LK+RP D + ++I F+ +

45

Sbjct: 17 LKEAAELLRNGEVLIIFPTETVYGIGADAYNEEACKKIFKLKERPADNPLIVHIHSFKQLE 76

Query: 61 KYSKNQPTYLKQLYDAFLPGPLTIIL-EASQEVPHWINSGLLSVGFRMPKHPVTLDMIAN 119  
+ ++ +L L F PGPLT+I + S+++P + + L +V RMP HPV L +I

Sbjct: 77 EIAEGYEPHLDLFL-KKFWPGPLTVIFRKKSEKIPPVVTDALPTVAVRMPAHPVALKLIEL 135

50

Query: 120 HG-PLIGPSANISGCDGRVFEIQKQFNHGV-LGIEDDKALTGVDSSTIIDLSGDRVKIL 177  
G P+ PSANISG S + + F +V L I+ G++STI+DL+ ++ +L

Sbjct: 136 FGHPIAAPSANISGRPSATNVKHVIEDFMGKVKLIIDAGDTPFGLESTIVDLTKKEPVL 195

Query: 178 RQGAIQTQEVLTATIPELI 195

55

R G + E L PEL+

Sbjct: 196 RPGPVEVERLKELFPELV 213

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6565> which encodes the amino acid sequence <SEQ ID 6566>. Analysis of this protein sequence reveals the following:

-2396-

Possible site: 46

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0282(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 127/196 (64%), Positives = 154/196 (77%)

Query: 1 MNDLGQILEDHGAVIMPTETVYGIFAKALSEEAVNHVYELKKRPRDKAMNLCDFETIL 60  
 M L I+E A+++PTETVYG+FAKAL E+AVN VY+LK+RPRDKAMNLN+ DF +IL  
 Sbjet: 11 MEYLASIIESGDALVLPETETVYGLFAKALDEKAVNAVYDLKQRPDKAMNLCNADFNSIL 70

15 Query: 61 KYSKNQPTYLKLQLYDAFLPGPLTIILEASQEVPHWINSGLLSVGFRMPKHPVTLDMIANH 120  
 +SK QP YLK+LY AFLPGPLTIIL+A+ +VP+WINSGL +VGFR+P HP+T +I  
 Sbjet: 71 AFSKEQPRYLKKLYQAFLPGPLTIILKANDQVPYWINSGLSTVGFRLPSPHPITAALIQKT 130

20 Query: 121 GPLIGPSANISGCDGRVFEIQQQFNHQVLGIEDDKALTGVDSTIIDLSGDRVKILRQG 180  
 GPLIGPSAN+SG SGRVF I + F+ QV G DD LTG DSTI+DLG+R ILRQG  
 Sbjet: 131 GPLIGPSANLSGKASGRVFDHIMQDFQVFGYADDPFLTGKDDSTIIDLSGERAVILRQG 190

25 Query: 181 AITQEVLTATIPELIF 196  
 AIT+E L A +PEL F  
 Sbjet: 191 AITKEELLANVPELRF 206

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**30 Example 2127**

A DNA sequence (GBSx2243) was identified in *S.agalactiae* <SEQ ID 6567> which encodes the amino acid sequence <SEQ ID 6568>. This protein is predicted to be protoporphyrinogen oxidase (hemK). Analysis of this protein sequence reveals the following:

Possible site: 50

35 &gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07493 GB:AP001519 protoporphyrinogen oxidase [Bacillus halodurans]  
 Identities = 94/236 (39%), Positives = 132/236 (55%), Gaps = 12/236 (5%)

45 Query: 49 DTDQQLMENIFQQLKKHRSP---QYITGKAYFRDLIFFVDERVLIIRPETEELVDLILSE 105  
 + D +L + + + L H S Q++ G F F VD+ VLIRPETEELV +L E  
 Sbjet: 46 ELDGELFQRLLEDLAHASGVVPVQHLIGVESFYGRQFQVDQHVLIIRPETEELVLAVLKE 105

50 Query: 106 -----NKVEDCSVLDDIGTGSGAIAISLKKERPSWDVLASDISVSALDLAKENANNCDAEV 160  
 K E+ ++LDIGTGSGAIA++L E +V A DIS AL +A +NA A V  
 Sbjet: 106 IRRQFKKEEITILDIGTGSGAIAVTLALEERTNVTAVDISRDALQVAADNARRLGANV 165

55 Query: 161 TFIESDV---FSNISGKFDIIVSNPPYISYNDKDEVGKNVLASEPHSALFADEEGLAIYR 217  
 I D+ F +FD+IVSNPPYI +KD + +V EP ALF +GL +YR  
 Sbjet: 166 QLIHGDLGEPFLKTGERFDVIVSNPPYIPTVEKDTLAVHVRDHEPALALFGGVDGLDVYR 225

60 Query: 218 KIIENSREYL-QPRGKLYFEIGYKQGDDLRLSLKRYFPNNRCRVLKDIFGKDRMVV 272  
 +++ + +G+ EIG QG D+ L++ +P VL D+ GKDR+V+  
 Sbjet: 226 RLMSQLPALTKKEKGMVALEIGAGQGMDVEKLMQTAYPKAAVDVLYDLNGKDRIVL 281

-2397-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6569> which encodes the amino acid sequence <SEQ ID 6570>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4324(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 174/274 (63%), Positives = 207/274 (75%)

Query: 1 MNYAQLIKHYGQLLEACGEEVENFIYVLKDLKQWSTTDYLLNQNSSVSDTDQQLMENIFQ 60  
MNYA LI+ Y LE E+ EN YV +++K+WS+ D L++QN +V+ D L+E+IF  
Sbjct: 1 MNYATLIRTYEDKLEQIDEDRENLAYVFREIKWSSLDMLIHQNQAVTPEDAVLLEHIFC 60

Query: 61 QLKKHRSPQYITGKAYFRDLIFVDERVLIIPRPETEELVDLILSENKVEDCSVLDIGTGS 120  
L +H SPQYITG AYFRDL VD+RVLIIPRPETEELVD+IL+EN +VLDIGTGS  
Sbjct: 61 SLSQHLSPQYITGNAYFRDLKLAVDKRVLIIPRPETEELVDMILAENLDAPLNVDIGTGS 120

Query: 121 GAIAISLKKERP+WDVLASDISVSALDLAKENANNCAEVTTFIESDVFSNISGKFDIIVS 180  
GAIAISLKKERP+W V ASDIS +ALDLAK NA+ ++TFIESDVFS IS FDIIVS  
Sbjct: 121 GAIAISLKKERP+WDVLASDISVSALDLAKENANNCAEVTTFIESDVFSNISGKFDIIVS 180

Query: 181 NPPYISYNDKDEVGKNVLAASEPHSALFADEEGLAIYRKIIENSREYLQPRGKLYFEIGYK 240  
NPPYISY DK+EV NVL SEPH ALFA E G AIYRKIIE + YL GKLYFEIGYK  
Sbjct: 181 NPPYISYEDKEEVSLNVLQSEPHLALFAKENGAIYRKIIEQADNYLTKEGKLYFEIGYK 240

Query: 241 QGDDLRSLLKRYFPNNRCRVLKDFGKDRMVVD 274  
Q + ++ +L+ YFP R + DIFGK+RMVV+D  
Sbjct: 241 QAEGIKDMLQAYFPQRHIRAVTDIFGKERMVVVD 274

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2128

A DNA sequence (GBSx2244) was identified in *S.galactiae* <SEQ ID 6571> which encodes the amino acid sequence <SEQ ID 6572>. This protein is predicted to be peptide chain release factor RF-1 (prfA).

Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3446(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15718 GB:Z99122 peptide chain release factor 1 [Bacillus subtilis]  
Identities = 211/351 (60%), Positives = 280/351 (79%), Gaps = 1/351 (0%)

Query: 5 DQLQAVEDRYEELGELSDPDVVSDFTRFMELSRREEASTRETVTAYREYKQVIQNISDAE 64  
D+L+++E+RYE+L ELLSDP+VV+D K+ E S+E++ +ETV YR+Y+ + ++DA+  
Sbjct: 3 DRLSKIEERYEKLNELSDPEVVDNPKKLREYSKEQSDIQETVDVYRQYRDASEQLADAK 62

Query: 65 EMIKASGDAAELEEMAKEELKESKAAKEEYERLKILLPKDPNDKNIILEIRGAAGGD 124  
M+++ DAE+ +M KEE+ E + E ERLK+LL+PKDPNDKNI+I+EIRGAAGG+

-2398-

Sbjct: 63 AMLEEKL-DAEMRDMVKKEISELQKETETLSERLKVLLIPKDPNDDKNVIMEIRGAAGGE 121

Query: 125 EAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLYESGAH 184  
EAALFAG+L MY +YAE QGW+ EVME++V G GG KE++ M++G YSKLKYE+GAH

5 Sbjct: 122 EAALFAGNLYRMYSRYAELQGWKTEVMEANVTGTGGYKEIIFMITGSGAYSKLKYENGAH 181

Query: 185 RVQRVPVTESQGRVHTSTATVLMPEVEEVEYEIDQKDLRVDIYHASGAGGQNVNKVATA 244  
RVQRVP TES GR+HTSTATV +PE EEVE +I +KD+RVD + +SG GGQ+VN +A

10 Sbjct: 182 RVQRVPETESGGRIHTSTATVACLPEAAEEVEVDIHEKDIRVDTFASSGPGGQSVNTTMSA 241

Query: 245 VRMVHIPTGIKVMQEERTQQKNRDKAMKIIIRARVADHFAQIAQDEQDAERKSTVGTGDR 304  
VR+ H+PTG+ V Q+E++Q KN++KAMK++RAR+ D F Q AQ E D RKS VG+GDR

Sbjct: 242 VRLTHLPTGVVSCQDEKSQIKNKEKAMKVLRRARIYDKFQQEAEYDQTRKSAVGSGDR 301

15 Query: 305 SERIRTYNFPQNRVTDHRIGLTIQKLDLILSGKMDEVIDALVMYDQTQKLE 355  
SERIRTYNFPQNRVTDHRIGLT+QKLD IL GK+DEV++AL++ DQ KL+

Sbjct: 302 SERIRTYNFPQNRVTDHRIGLTIQKLDQILEGKLDDEVVEALIVEDQASKLQ 352

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6573> which encodes the amino acid  
20 sequence <SEQ ID 6574>. Analysis of this protein sequence reveals the following:

Possible site: 28  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.3446(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 349/358 (97%), Positives = 354/358 (98%)

Query: 1 MNIYDQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREEASTRET V TAYREYKQVIQNI 60  
MNIYDQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREE +TRET V TAYREYKQVIQ I

35 Sbjct: 1 MNIYDQLQAVEDRYEELGELLSDPDVVS DTKRFMELSREE T N TRET V TAYREYKQVIQTI 60

Query: 61 SDAEEMIKDASGDAELEEMAKEELKESKAAKEEYERL KILL L L P K D P N D D K N I I L E I R G A 120  
SDAEEMIKDASG D ELEEMAKEELKESKAAKEEYEE+L KILL L L P K D P N D D K N I I L E I R G A

Sbjct: 61 SDAEEMIKDASG D P E L E E M A K E E L K E S K A A K E E Y E E K L K I L L L P K D P N D D K N I I L E I R G A 120

40 Query: 121 AGGDEAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLYE 180  
AGGDEAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLYE

Sbjct: 121 AGGDEAALFAGDLLTMYQKYAETQGWRFEVMESSVNGVGGIKEVVAMVSGQSVYSKLYE 180

45 Query: 181 SGAHRVQRPVPTESQGRVHTSTATVLMPEVEEVEYEIDQKDLRVDIYHASGAGGQNVNK 240  
SGAHRVQRPVPTESQGRVHTSTATVLMPEVEEVEY+ID KDLRVDIYHASGAGGQNVNK

Sbjct: 181 SGAHRVQRPVPTESQGRVHTSTATVLMPEVEEVEYDIDPKDLRVDIYHASGAGGQNVNK 240

Query: 241 VATAVRMVHIPTGIKVMQEERTQQKNRDKAMKIIIRARVADHFAQIAQDEQDAERKSTVG 300  
VATAVRMVHIPTGIKVMQEERTQQKNRDKAMKIIIRARVADHFAQIAQDEQDAERKSTVG

50 Sbjct: 241 VATAVRMVHIPTGIKVMQEERTQQKNRDKAMKIIIRARVADHFAQIAQDEQDAERKSTVG 300

Query: 301 TGDRSERIRTYNFPQNRVTDHRIGLTIQKLDLILSGKMDEVIDALVMYDQTQKLEALN 358  
TGDRSERIRTYNFPQNRVTDHRIGLTIQKLDLILSGKMDEVIDALVMYDQT+KLE+LN

55 Sbjct: 301 TGDRSERIRTYNFPQNRVTDHRIGLTIQKLDLILSGKMDEVIDALVMYDQTKKLES LN 358

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
vaccines or diagnostics.



-2399-

**Example 2129**

A DNA sequence (GBSx2245) was identified in *S.agalactiae* <SEQ ID 6575> which encodes the amino acid sequence <SEQ ID 6576>. This protein is predicted to be thymidine kinase (tdk). Analysis of this protein sequence reveals the following:

```

5   Possible site: 39
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2244 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9841> which encodes amino acid sequence <SEQ ID 9842> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAB02289 GB:L40415 thymidine kinase [Streptococcus gordonii]
    Identities = 158/189 (83%), Positives = 175/189 (91%)

20   Query: 1  MAQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDEFGVSSRIGMRREAV 60
      MAQLYYKYGTMSGKTIEILKVAHNYEEQGK VVIMTSA+DTRD G VSSRIGM+R+A+
      Sbjct: 1  MAQLYYKYGTMSGKTIEILKVAHNYEEQGKGVVIMTSAVDTRDGVGYVSSRIGMKRQAM 60

    Query: 61  PISDDMDIFSYIQNLPQKPYCVLIDECQFLSKKNVYDLARVVDDLDVPVMAFGLKNDFQN 120
      I DD DI YI+NLP+KPYC+LIDE QFL + +VYDLARVVD+LDVPVMAFGLKNDF+N
25   Sbjct: 61  AIEDTDILGYIKNLPKPYCILIDEAQFLKRHHVYDLARVVDELDPVMAFGLKNDFRN 120

    Query: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
      LFEGSKHLLLLADKI+EIKTICQYCS+KATMVLRT++GKPVY+G+QIQIGGNETYIPVC
30   Sbjct: 121 ELFEGSKHLLLLADKIEIKTICQYCSRKATMVLRTDHGKPVYDGEQIQIGGNETYIPVC 180

    Query: 181 RKHYFNPDI 189
      RKHYF PDI
30   Sbjct: 181 RKHYFKPDI 189

```

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6577> which encodes the amino acid sequence <SEQ ID 6578>. Analysis of this protein sequence reveals the following:

```

    Possible site: 39
    >>> Seems to have no N-terminal signal sequence

40   ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2244 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

45 An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 174/189 (92%), Positives = 184/189 (97%)

50   Query: 1  MAQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDEFGVSSRIGMRREAV 60
      +AQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRD FG+VSSRIGMRREA+
      Sbjct: 1  LAQLYYKYGTMSGKTIEILKVAHNYEEQGKPVVIMTSALDTRDGFIVSSRIGMRREAI 60

    Query: 61  PISDDMDIFSYIQNLPQKPYCVLIDECQFLSKKNVYDLARVVDDLDVPVMAFGLKNDFQN 120
      PIS+DMDIF++I L +KPYCVLIDE QFLSK+NVYDLARVVD+L+VPVMAFGLKNDFQN
55   Sbjct: 61  PISNDMDIFTFIAQLEEKPYCVLIDESQFLSKQNVYDLARVVDELNVPVMAFGLKNDFQN 120

    Query: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180
      NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC
      Sbjct: 121 NLFEGSKHLLLLADKIDEIKTICQYCSKKATMVLRTENGKPVYEGDQIQIGGNETYIPVC 180

```

-2400-

Query: 181 RKHYFNPD1 189  
 RKHYFNPD1  
 Sbjct: 181 RKHYFNPD1 189

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2130

A DNA sequence (GBSx2246) was identified in *S.agalactiae* <SEQ ID 6579> which encodes the amino acid sequence <SEQ ID 6580>. Analysis of this protein sequence reveals the following:

10 Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 15 bacterial cytoplasm --- Certainty=0.3995(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAA26046 GB:M95650 4-oxalocrotonate tautomerase [Plasmid pWW0]  
 Identities = 27/60 (45%), Positives = 36/60 (60%)

Query: 1 MPFVKIDLFEGRSQEQKNELAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGE LKKK 60  
 MP +I + EGRS EQK L REV+E +SR AP ++ V I +M +G + GEL K  
 25 Sbjct: 1 MP1AQIHILEGRSDEQKETLIREVSEAISRSILDAPLTSVRVIITEMAKGHFGIGGELASK 60

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6581> which encodes the amino acid sequence <SEQ ID 6582>. Analysis of this protein sequence reveals the following:

30 Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4128(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 56/60 (93%), Positives = 59/60 (98%)

40 Query: 1 MPFVKIDLFEGRSQEQKNELAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGE LKKK 60  
 MPFV IDLFEGRSQEQKN+LAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGE+K+K  
 Sbjct: 1 MPFVTIDLFEGRSQEQKNQLAREVTEVVSRIAKAPKENIHVFINDMPEGTYYPQGE MKQK 60

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 45 Example 2131

A DNA sequence (GBSx2247) was identified in *S.agalactiae* <SEQ ID 6583> which encodes the amino acid sequence <SEQ ID 6584>. Analysis of this protein sequence reveals the following:

50 Possible site: 57  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2154(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2401-

A related GBS nucleic acid sequence <SEQ ID 9843> which encodes amino acid sequence <SEQ ID 9844> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5   >GP:AAC65759 GB:AE001250 conserved hypothetical protein [Treponema
      pallidum]
      Identities = 103/317 (32%), Positives = 163/317 (50%), Gaps = 15/317 (4%)

10  Query: 7   QLSHSLRLMGTTIDIQINSKNAQKQIR---EVIELLELYKNRFSANDEFNSELMAINNNA 62
      + S + ++GT +++ SK ++ EV LL+ + SAN +S L A+N A
      Sbjct: 31 EYSRAELVIGTLCRVRVYSKRPAAEVHAALAEVFTLLQQQEMVLSANRDDSALAALNAQA 90

      Query: 63 GIKPIQVHPDLFELITIGKEHSLARPSNLNIAIGPLVQTRIGFSDAKLPSPSEISEAMI 122
      G P+ V L+ L+ + N A+G V+ W IGF A +P P + EA+
15  Sbjct: 91 GSAPVVVDRSLYALLERALFFAEKSGGAFNPALGAXVKLWNIGFDRAAVDPDPALKEALT 150

      Query: 123 LSDPTHILLDSN----KQSVFLNQIGMKIDLGALAKGYIADKIMTYLKNEMIDSAIINL 177
      D + L + +V L Q GM++DLGA+AKG++ADKI+ L +DSA+++L
20  Sbjct: 151 RCDFRQVHLRAGVSVGAPHTVQLAQAGMQLDLGAIAGFLADKIVQLLTAHALDSALVDL 210

      Query: 178 GGNV----LVHGDNPNRSEGY--WVIGIQHPKKRGNIGTVKIKNQSVVTSGETYERRLI 231
      GGN+ L +GD + + W +GI+ P K V +++ SVVTSG YER
      Sbjct: 211 GGNIFALGLKYGDVRSAAQRLEWNVGIRDPHGTGQKPALVVSVRDCSVVTSGAYERFFE 270

25  Query: 232 IDKEYHHIFDRQTGYPIQTEMASISIVSKQSVDCIEWTTRLFGLSIKEALDILNAVSYI 291
      D YHHI D TG+P T++ S+SI + +S D + T F L +++ +L +
      Sbjct: 271 RDGVRYHHIIDPVTGFPAHTDVSISIFAPRSTDADALATACFVLGYEKSCALLREFPGV 330

      Query: 292 EGIITTKDDRIYLSDDL 308
      + + I D R+ S G+
30  Sbjct: 331 DALFIFPDKRVRASAGI 347

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6585> which encodes the amino acid sequence <SEQ ID 6586>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 52
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1020(Affirmative) < succ>
40  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 182/310 (58%), Positives = 232/310 (74%)

45  Query: 8   LSHSLRLMGTTIDIQINSKNAQKQIREVIELLELYKNRFSANDEFNSELMAINNAGIKPI 67
      ++ L+LMGT IDIQI S A +Q+ VI+LL YKNRFSAND NSELMAIN AG+KP+
      Sbjct: 3   VTQQLKLMGTVIDIQIESDKACQQLSRVIDLLTYTKNRFSANDSNSELMAINQAAGVKPV 62

      Query: 68. QVHPDLFELITIGKEHSLARPSNLNIAIGPLVQTRIGFSDAKLPSPSEISEAMILSDPT 127
      VH DLF LI IGK HSL+ PSNLNIAIGPLVQ WRIGF DA++PS + IS+ + L+DP
50  Sbjct: 63 SVHSDLFNLIQIGKAHSLSTPSNLNIAIGPLVQAWRIGFEDARVP SHNLISQQLALTDPR 122

      Query: 128 HILLDSNKSQSVFLNQIGMKIDLGALAKGYIADKIMTYLKNEMIDSAIINLGGNVLVHGDN 187
      +L+D KQ+VFL Q+GM +DLGALAKGYI DKIM YL + IDSA+INLGGNV VH G N
55  Sbjct: 123 QVLIDDKKQTVFLQQVGMALDLGALAKGYITDKIMAYLIEDGIDSALINLGGNVVRVHGPN 182

      Query: 188 PNRSEGYWVIGIQHPKKRGNIGTVKIKNQSVVTSGETYERRLIIDKEYHHIFDRQTGY 247
      P + + IGIQ P KRG+++G +K+ N SVVTSG YER+ K+YHHI DRQTGY
60  Sbjct: 183 PKSPDKTFRIGIQKPDAKRGQHLGVKVNHSVVTSGIYERQFTSKGKQYHHILDRQTGY 242

      Query: 248 PIQTEMASISIVSKQSVDCIEWTTRLFGLSIKEALDILNAVSYIEGIITTKDDRIYLSDG 307

```

-2402-

PI+T+M S++I++ S C+IWTRLFGL + +LN IEG+++T+ + +S+G  
 Sbjct: 243 PIETDMLSLTIMAPSSFYCDIWTRLFGLDSSMIITLLNTFDNIEGLLVTRKHHVLMNSG 302

Query: 308 LKHHFQLFYH 317  
 L+H+FQ +YH  
 Sbjct: 303 LRHYFQPYH 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 2132

A DNA sequence (GBSx2248) was identified in *S. agalactiae* <SEQ ID 6587> which encodes the amino acid sequence <SEQ ID 6588>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.0966(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG18632 GB:AY007504 unknown [Streptococcus mitis]  
 Identities = 92/160 (57%), Positives = 119/160 (73%), Gaps = 1/160 (0%)

Query: 1 MKLIGIVGTNSNKSTNRQLLQYMQHFADKAEIELEIEVKDLPLFNKPADKNVPQVILDIA 60  
 MKL+ IVGTNSN+STNR+LL++MQ+HF+DKA+IE++E+K LP FN+P D+ P + +  
 Sbjct: 1 MKLVAIVGTNSNRSTNRKLLKFMQKHFSKADIEVLEIKQLPAFNEPEDEQAPAEVQAQS 60

Query: 61 AKIEETDGVIIIGTPEYDHSIPSALMSVLAWLSYGIYPLLNKPMITGASYGTLGSSRAQL 120  
 KI DGVII TPEYDH+IP+ L S L W++Y L+NKP MI GAS G LG+SRAQ  
 Sbjct: 61 EKILAADGVIIISTPEYDHTIPAPLASALEWIAYTSTRALINKPTMIVGASLGLLGTSRAQA 120

Query: 121 QLQILNAPELKASVLP-DEFLLSHSLQAFDKDGNLHDIE 159  
 LRQIL+APELKA V+P EF L HS Q D + +L++ E  
 Sbjct: 121 HLRQILDAPELKARVMPGTEFFLGHSEQVLDDECHLNPE 160

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6589> which encodes the amino acid sequence <SEQ ID 6590>. Analysis of this protein sequence reveals the following:

Possible site: 24  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB62679 GB:AL133422 putative secreted protein. [Streptomyces coelicolor A3(2)]  
 Identities = 68/192 (35%), Positives = 94/192 (48%), Gaps = 25/192 (13%)

Query: 4 ILFIVGSLREGSFNHQLAAQAQK-ALEHQAVVSYLNWKDVPVLNQDIEANAPLPVVDA-- 60  
 IL +VGSRLR GS N QLA A + A E V + ++P N+DI+ +P A  
 Sbjct: 5 ILALVGSLRAGSHNRQLAEAAVRFAPEGAIEVQLFEGLAIEIPFYNEDIDVEGVSPPAAAKL 64

Query: 61 RQAVQSADAIWIFTPVYNFSIPGSVKNNLLDWLSRALDLSDPGPSAIGGKVVTVSSVANG 120  
 R+A Q A A +F+P YN +IP +KN +DWLSR P G A GK V V A G  
 Sbjct: 65 REAAQGAQAFLLFSPEYNGTIPAVLKNAIDWLSR-----PYGAGFTGKPVAVVGTAFG 118

-2403-

Query: 121 GHDQVFDQFKA-----LLPFIRTSVAGEFTK-ATVNP--DAWGTGRLEISKETKA 167  
 + V+ Q +A ++ I+ S+ G T+ A +P DA +L E A  
 Sbjet: 119 QYGGVWAQDEARKAVGIAGGKVIEDIKLSIPGSVTRFAETHPADDAEVAAQL---TEVVA 175

Query: 168 NLLSQAEALLAA 179  
 L A+ +AA  
 Sbjet: 176 RLHGHADAIAA 187

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 28/90 (31%), Positives = 49/90 (54%)

Query: 3 LIGIVGTNSNKSINRQLLOQMQQHFADKAEIELIEVKDLPLFNKPADKNVPQVILDIAAK 62  
 ++ IVG+ S N QL Q+ +A + + KD+P+ N+ + N P ++D  
 Sbjet: 4 ILFTIVGSLREGSFNHLAAQAQKALEHQAVVSYNWIKDVPVLNQDIEANAPLPVVDARQA 63

Query: 63 IEETDGVIIIGTPEYDHSIPSALMSVLAWLS 92  
 ++ D + I TP Y+ SIP ++ ++L WLS  
 Sbjet: 64 VQSADAIWIFTPVYNFSIPGSVKNLLDWLS 93

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2133

A DNA sequence (GBSx2249) was identified in *S.agalactiae* <SEQ ID 6591> which encodes the amino acid sequence <SEQ ID 6592>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1160(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2134

A DNA sequence (GBSx2250) was identified in *S.agalactiae* <SEQ ID 6593> which encodes the amino acid sequence <SEQ ID 6594>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2132(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG18632 GB:AY007504 unknown [Streptococcus mitis]  
 Identities = 80/162 (49%), Positives = 112/162 (68%)

Query: 1 MKFVGIVGSNAEQSYNRMLEFIRKNFKTKFELEVLEIDDIPMFNQDQNWEEFQLRLLN 60

-2404-

```

          MK V IVG+N+ +S NR LL+F++K+F K ++EVLEI +P FN+ ++ + +++ +
Sbjct: 1  MKLVAIVGTNSNRSTNRKLLKFMQKHFSKADIEVLEIKQLPAFNEPEDEQAPAEVQAFS 60

5  Query: 61  NKITRADGVIIATPEHNHTITAALKSVLEWLSFAVHPLENKPMIVGASYDQGTSRAQI 120
          KI  ADGVII+TPE++HTI A L S LEW+++ L NKP MIVGAS GTSRAQ
Sbjct: 61  EKILAADGVIIISTPEYDHTIPAPLASALEWIAYTSTRALINKPTMIVGASLGLLGTSRAQA 120

          Query: 121 HLRKILDAPGVNAYTLPGNEFLLGKAKEAFDDNGNIINPGTV 162
          HLR+ILDAP + A +PG EF LG +++ DD ++ NP V
10  Sbjct: 121 HLRQILDAPELKARVMPGTEFFLGHSEQVLDDECHLNNPEKV 162

```

There is also homology to SEQ ID 6596.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 15 Example 2135

A DNA sequence (GBSx2251) was identified in *S.agalactiae* <SEQ ID 6597> which encodes the amino acid sequence <SEQ ID 6598>. Analysis of this protein sequence reveals the following:

```

Possible site: 21
>>> Seems to have no N-terminal signal sequence
20  INTEGRAL    Likelihood = -7.32    Transmembrane    13 - 29 ( 11 - 29)

----- Final Results -----
          bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
25  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2136

A DNA sequence (GBSx2252) was identified in *S.agalactiae* <SEQ ID 6599> which encodes the amino acid sequence <SEQ ID 6600>. This protein is predicted to be potential nitrite transporter. Analysis of this protein sequence reveals the following:

```

35  Possible site: 42
>>> Seems to have no N-terminal signal sequence
          INTEGRAL    Likelihood = -9.92    Transmembrane    61 - 77 ( 54 - 82)
          INTEGRAL    Likelihood = -5.57    Transmembrane    106 - 122 ( 103 - 126)
          INTEGRAL    Likelihood = -5.15    Transmembrane    160 - 176 ( 159 - 177)
40  INTEGRAL    Likelihood = -4.09    Transmembrane    180 - 196 ( 179 - 199)
          INTEGRAL    Likelihood = -1.01    Transmembrane    233 - 249 ( 233 - 249)

----- Final Results -----
          bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
45  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB15832 GB:Z99123 alternate gene name: ipa-48r~similar to
50  nitrite transporter [Bacillus subtilis]
          Identities = 82/253 (32%), Positives = 119/253 (46%), Gaps = 10/253 (3%)

          Query: 6  EKIAYNCAKKEALYKESLGRYALRSLAGAYLTMTSTAAGIVAADTIGK-ISPALSGFVF- 63

```

-2405-

+K+      KK+ ++   S   RY   LRS+LA   ++      GI   AA   G      A   S   F   F  
 Sbjct: 7    QKVEQYALKKQNI FASSKIRYVLRSLASIFIGF----GITAASKTGSYFFMADSPFAFP 62

Query: 64    --AFIFSFLIYVLIFNGELATSNMLYLTAGAYNKNISWKKAMTILIYCTFFNLVGACIL 121  
           A   F    ++ +      G+L   T   N   Y   T   A   K   ISW+   + + +      NL+GA +  
 Sbjct: 63    AA AVTFGAAILMIAYGGDLFTGNTFYFTYTALRKKISWRDTLYLWMSYAGNLIGAILF 122

Query: 122    AWLFNQSYSFQHLTND SFLGHVVAKKLGKPSGAFLEGIIANMFVNAILAYMLLKEESA 181  
           A   L   + +   F+   +   SFL   H+   K+   P+S   F   G++   N   V   LA      M   LK   EA  
 Sbjct: 123    AILISATGLFEPSVHSFLIHLAEHKMEPPASELFFRGMLCNWLVLCAFFIPMSLKGEGA 182

Query: 182    KMTVILSAIFMFVFLSNEHLIANFASFMLAASFHIEHIKGFLLNIIRQWTLVFFGNWIG 241  
           K+   ++   +F   F      EH   IAN   +F   ++      IEH      TL+   +R      V   GN  
 Sbjct: 183    KLFTMMLFVFCFFISGFHESIANMCTFAISLL--IEHPDVTLMGAVRNLIPTVLGNLTA 240

Query: 242    GGVFIGLAYAWLN 254  
           G   V   +G   Y   LN  
 Sbjct: 241    GIVMMGWMYYTLN 253

20    A related DNA sequence was identified in *Spyogenes* <SEQ ID 6601> which encodes the amino acid sequence <SEQ ID 6602>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

INTEGRAL    Likelihood = -9.77    Transmembrane    142 - 158 ( 139 - 171)  
 INTEGRAL    Likelihood = -9.34    Transmembrane    95 - 111 ( 89 - 119)  
 INTEGRAL    Likelihood = -2.02    Transmembrane    61 - 77 ( 61 - 79)  
 INTEGRAL    Likelihood = -1.12    Transmembrane    261 - 277 ( 261 - 279)  
 INTEGRAL    Likelihood = -0.53    Transmembrane    191 - 207 ( 191 - 207)

----- Final Results -----

bacterial membrane --- Certainty=0.4906(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35    The protein has homology with the following sequences in the databases:

>GP:AAB80864 GB:U93874 formate dehydrogenase [Bacillus subtilis]  
 Identities = 133/258 (51%), Positives = 181/258 (69%)

Query: 36    KTPEQILEATIHIGEHKVTKTFLAKSILGFIGGAMISLGYLLYVRIAASGLETFGAFSSI 95  
           + P++I   EA   I   G   K+      +   +LGF+GGA   I+LGYLL   +R+      + +G+   SS+  
 Sbjct: 4    RKPDEIAEAAIEAGMKKIKLPLPSLLVLGFLGGAFFIALGYLLDIRVIGDLPKEWGLSSSL 63

Query: 96    VGACAFFIGLIILMAGGELITGNMMAVSAALLAKKIKFSELAKNWLIITLFNVIGAVFV 155  
           +GA   FP+GLI++++AG   ELITGNMM+V+   AL   ++KI      ELA   NW   I+T+   N+IGA+FV  
 Sbjct: 64    IGA AVFPVGLILVVLGAELITGNMMSVAMALFSRKISVKELAINWGIVTIMNLIGALFV 123

Query: 156    AFVFGHFLGLTSAGIFKEEVIEVAHAKIAASPLQALVSGIGCNWVFLALWLCYANDAA 215  
           A+   FGH   +GLT   G   +   E+   I   VA   K+   S   +   L+S   IGCNW   V   LA+WL   +GA   DAA  
 Sbjct: 124    AYFFGHVLGLTETGPYLEKTI A VAQGLDMSFGKVLISAICNWLVLAVWLSFGAQDAA 183

Query: 216    GKFLGTWFPVMTFVALGFQHSVANAFVIPAAIFEGGATWLDFTNFI FVYSGNIIGGAIF 275  
           GK   LG   WFP+M   FVA+GFQH   VAN   FVIPAAIF   G   TW   F+   N   I   +   GN+IGGA+F  
 Sbjct: 184    GKILGIWFPIMAFVAIGFQHVANMFVIPAAIFAGSFTWGFIGNIIPAFIGNVIGGAVF 243

Query: 276    VSFLYFKVYYHPQKSKTQ 293  
           V   +YF   Y+      +S+   +  
 Sbjct: 244    VGLIYFIAYHKKDRSRKE 261

An alignment of the GAS and GBS proteins is shown below.

Identities = 69/240 (28%), Positives = 101/240 (41%), Gaps = 18/240 (7%)

Query: 15    KEALYKESLGRYALRSLAGAYLTMTAAAGIVAADTIGKISPALSGFVFVAFIFSFLIYV 74  
           K   L   K   LG      +   G   L   +   AA      +T   G      A   S   V   A   F   GLI   +  
 Sbjct: 55    KTFLAKSILGFIGGAMISLGYLLYVRIAAS--GLETFG----AFSSIVGACAFFIGLIIL 108

-2406-

Query: 75 LIFNGELATSNMLYLTAGAYNKNISWKKAMTILIYCTFFNLVGACILAWLFNQSYSFQHL 134  
 L+ GEL T NM+ ++A K I + + + T FN++GA +A++F F L  
 Sbjct: 109 LMAGGELITGNMMAVSAALLAKKIKFSELAKNWLITLTFNVIGAVFVAFVFGH---FLGL 165

Query: 135 TNDNFLGHVVAK---KLKGPSSGAFLEGIIANMFVNLAAILAYMLLKEESAKMTVILSAI 190  
 T+ V + K+ A + GI N FV LA+ + + K +  
 Sbjct: 166 TSAGIFKEEVIEVAHAKIAASPLQALVSGIGCNWVGLALWLCYGANDAAGKFLGTWFPV 225

Query: 191 FMFVFLSNEHLIANFASFMLAASHIEHIKGFLLNIIRQWTLVFFGNWIGGGVFIGLAY 250  
 FV L +H +AN A F G T L+ + + V+ GN IGG +F+ Y  
 Sbjct: 226 MTFVALGFQHSVANAFVIPAAIFE-----GGATWLDFTNFIIVYSGNIIGGAIFVSFLY 280

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2137

A DNA sequence (GBSx2253) was identified in *S.agalactiae* <SEQ ID 6603> which encodes the amino acid sequence <SEQ ID 6604>. Analysis of this protein sequence reveals the following:

Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1342(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2138

A DNA sequence (GBSx2254) was identified in *S.agalactiae* <SEQ ID 6605> which encodes the amino acid sequence <SEQ ID 6606>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.22 Transmembrane 44 - 60 ( 44 - 60)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1086(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.



-2407-

**Example 2139**

A DNA sequence (GBSx2255) was identified in *S.agalactiae* <SEQ ID 6607> which encodes the amino acid sequence <SEQ ID 6608>. This protein is predicted to be xanthine permease (pbuX). Analysis of this protein sequence reveals the following:

```

5   Possible site: 23
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.91    Transmembrane 160 - 176 ( 156 - 188)
      INTEGRAL    Likelihood = -6.48    Transmembrane 184 - 200 ( 179 - 211)
      INTEGRAL    Likelihood = -6.21    Transmembrane 101 - 117 ( 96 - 121)
10   INTEGRAL    Likelihood = -4.04    Transmembrane 309 - 325 ( 306 - 332)
      INTEGRAL    Likelihood = -3.98    Transmembrane 334 - 350 ( 331 - 353)
      INTEGRAL    Likelihood = -3.88    Transmembrane 400 - 416 ( 396 - 420)
      INTEGRAL    Likelihood = -3.45    Transmembrane 19 - 35 ( 18 - 38)
      INTEGRAL    Likelihood = -2.81    Transmembrane 127 - 143 ( 127 - 146)
15   INTEGRAL    Likelihood = -2.71    Transmembrane 228 - 244 ( 227 - 249)
      INTEGRAL    Likelihood = -2.02    Transmembrane 47 - 63 ( 47 - 63)
      INTEGRAL    Likelihood = -1.97    Transmembrane 75 - 91 ( 73 - 92)
      INTEGRAL    Likelihood = -0.85    Transmembrane 368 - 384 ( 368 - 384)

20   ----- Final Results -----
      bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25   The protein has homology with the following sequences in the GENPEPT database.
      >GP:CAB14123 GB:Z99115 xanthine permease [Bacillus subtilis]
      Identities = 213/412 (51%), Positives = 292/412 (70%), Gaps = 5/412 (1%)

30   Query: 14  LGLQHLLAMYAGSILVPIIMIASALGYNAKQLTYLIATDIFMCGIATLLQLRLSKHFGVGL 73
      LG+QH+LAMYAG+I+VP+++ A+G +QLTYL++ DIFMCG+ATLLQ+ ++ FG+GL
      Sbjct: 11  LGIQHVLAMYAGAIVVPLIVGKAMGLTVEQLTYLVSIDIFMCGVATLLQVWSNRFFGIGL 70

      Query: 74  PVVLGCAPQSVAPLSIIGAQQGSGYMGFALIASGIYVVLVAGIFSKVANFFPPIVTSVI 133
      PVVLGC F +V+P+ IG++ G ++G++IASGI V+L++ F K+ +FFPP+VTGGSV+
35   Sbjct: 71  PVVLGCTFTAVSPMIAIGSEYGVSTVYGSIIASGILVILISFFFFGKLVSFFPPVVTGSSV 130

      Query: 134 TTIGLTLIPVAMGNMGD--NAKEPSLQSLTSLVTIGVLLINIFAKGFLKSISILIGL 190
      T IG+TL+PVAM NM +A L +L L+ + +++L+ F KGF+KS+SILIG+
40   Sbjct: 131 TIIGITLMPVAMNMMAGGEGSADFGDLNLAFAFTVLSIIIVLLYRFTKGFIKSVSILIGI 190

      Query: 191 ISGTILAAFMGLVDASVVADAPLVHIPKPFYFGAPRFEFTSILMMCIATVSMVESTGVY 250
      + GT +A FMG V V+DA +V + +PFYFGAP F I+ M I+A VS+VESTGVY
      Sbjct: 191 LIGTFIAYFMGKVQFDNVSDAAVQMIQPFYFGAPSFHAAPIITMSIVAIVSLVESTGVY 250

45   Query: 251 LALSDITNDKLDKRLRNGYRSEGLAVLLGGLFNTFFPYTGFSQNVGLVQISGIRTRKPIY 310
      AL D+TN +L L GYR+EGLAVLLGG+FN FPYT FSQNVGLVQ++GI+ I
      Sbjct: 251 FALGDLTNRRLTEIDLKGYRAEGLAVLLGGIFNAPFYTAFSQNVGLVQLTGIKKNAVIV 310

      Query: 311 FTALFLVILGLLPKFGAMAQMIQSPVLGGAMLVLFMGMVALQGMKMLNQVDFEHNEHNFII 370
      T + L+ GL PK A +IFS VLGGAM+ +FGMV G+KML+++DF E N +I
50   Sbjct: 311 VTGVILMAFGLFPKIAAFTTIIPSAVLGGAMVAMFGMVIAYGIKMLSRIDFAKQE-NLLI 369

      Query: 371 AAVSIAAGVGFGNGT-NLFISLPNTLQMFNTNGIVISTLTAVVLNIINGLPK 421
      A S+ G+G ++F LP+ L + TNGIV + TAVVLNI+ N K
55   Sbjct: 370 VACSVGLGLGVTVPDPDFKQLPSALTLLTNGIVAGSFTAVVLNIIVNVFSK 421

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6609> which encodes the amino acid sequence <SEQ ID 6610>. Analysis of this protein sequence reveals the following:

```

60   Possible site: 29
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -7.32    Transmembrane 160 - 176 ( 158 - 181)
      INTEGRAL    Likelihood = -6.37    Transmembrane 103 - 119 ( 98 - 124)

```

-2408-

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -5.84 | Transmembrane | 130 - 146 ( 126 - 152) |
| INTEGRAL | Likelihood = -5.68 | Transmembrane | 187 - 203 ( 182 - 207) |
| INTEGRAL | Likelihood = -3.98 | Transmembrane | 337 - 353 ( 334 - 356) |
| INTEGRAL | Likelihood = -3.82 | Transmembrane | 232 - 248 ( 225 - 252) |
| INTEGRAL | Likelihood = -3.35 | Transmembrane | 403 - 419 ( 399 - 421) |
| INTEGRAL | Likelihood = -2.50 | Transmembrane | 22 - 38 ( 21 - 41)     |
| INTEGRAL | Likelihood = -2.07 | Transmembrane | 312 - 328 ( 312 - 328) |
| INTEGRAL | Likelihood = -1.97 | Transmembrane | 78 - 94 ( 76 - 95)     |

10 ----- Final Results -----  
                   bacterial membrane --- Certainty=0.3930(Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 The protein has homology with the following sequences in the databases:

>GP:CAB15234 GB:Z99120 similar to purine permease [Bacillus subtilis]  
 Identities = 216/421 (51%), Positives = 302/421 (71%), Gaps = 5/421 (1%)

Query: 6 KQEHSHSQSAVLGLQHVLSMYAGSILVPIMIAGALGYSARELTYLISTDIFMCGVATFLQ 65  
 K++H+ Q +LGLQH+L+MYAG+ILVP+++ A+G +A +LTYLI+ D+FMCG AT LQ  
 Sbjet: 2 KEQHNAQLMMLGLQHMLAMYAGAILVPLVGAATGLNAGQLTYLIAIDLFMCGAATLLQ 61

Query: 66 LKLTKHTGVGLPVVLGCAFSQSVAPLSIIGAQQGSGAMFGALIASGIYVILVAGIFSKIAR 125  
 L ++ G+GLPVVLGC F +V P+ IG+ G A++GA+IA+G+ V+L AG F K+ R  
 Sbjet: 62 LWRNRYFGIGLPPVVLGCTFTAVGPMISIGSTYGVPAIYGAILAAGLIVVLAAGFFGKLVR 121

Query: 126 FFPPIVTGSVITVIGLSLVGVAMGNM--GDNVKE-PTAQSMMLSLTTIVIIILLVQKFTKG 182  
 FFPP+VTGSV+ +IG+SL+ AM N+ G+ KE + +++L ILL+ F KG  
 Sbjet: 122 FFPPVVTGSVVMIIIGISLIPTAMNLAGGEGSKEFGSLDNVLLGFGVTAFILLFYFFKG 181

Query: 183 FVKISILIGLVAGTLVSAMMGLVDTPVVEASWIHVPTPFYFGMPTFEITSIVMMCIIA 242  
 F++SI+IL+GL+AGT + MG VD + V+EASW+HVP+ FYFG PTFE+ ++V M ++A  
 Sbjet: 182 FIRSIAILLGLIAGTAAAYFMGKVDFSEVLEASWLHVPSLFYFGPPTFELPAVVTMLLVA 241

Query: 243 TVSMVESTGVYLALSDLTNDQLDEKRLRNGYRSEGIAGVFLGGLFNTFPYTGFSONVGLVQ 302  
 VS+VESTGVY AL+D+TN +L EK L GYR+EG+A+ LGGLFN FPYT FSQNVG+VQ  
 Sbjet: 242 IVSLVESTGVYFALADITNRRLSEKDLKGYRAEGLAILLGGLFNAPFYTAFSQNVGIVQ 301

Query: 303 ISGIKTRRPIYYAAGILVIGLLPKFRAMAQMIPSPVLGGAMLVLFGMVALQGMQMLNRV 362  
 +S +K+ I ILV IGL+PK A+ +IP+PVLGGAM+V+FGMV G++ML+ V  
 Sbjet: 302 LSKMKSVMNVIAITGIILVAIGLVPKAAALTTVIPTVVLGGAMIVFMGMVISYGIKMLSSV 361

Query: 363 DFQKNEYNFIIAAVSISAGLGFNGT-NLFASLPETAQMFLTNGIVIAITLTSVVLNLVLNGK 422  
 D ++ N +I A S+S GLG LF+SL A + +GIVI +LT++ L+ K  
 Sbjet: 362 DLD-SQGNLLIIASSVSLGLGATTVPALFSSSLGAASVLGSGIVIGSLTAIALHAFQTK 421

An alignment of the GAS and GBS proteins is shown below.

Identities = 328/416 (78%), Positives = 380/416 (90%)

Query: 7 SNSQAALLGLQHLLAMYAGSILVPIMIASALGYNAKQLTYLIATDIFMCGIATLLQLRLS 66  
 S+SQ+A+LGLQH+L+MYAGSILVPIMIA ALGY+A++LTYLI+TDIFMCG+AT LQL+L+  
 Sbjet: 10 SHSQSAVLGLQHVLSMYAGSILVPIMIAGALGYSARELTYLISTDIFMCGVATFLQLKLT 69

Query: 67 KHFGVGLPVVLGCAFSQSVAPLSIIGAQQGSGYMFAGALIASGIYVVLVAGIFSKVANFFPP 126  
 KH GVGLPVVLGCAFSQSVAPLSIIGAQQGSG MFGALIASGIYV+LVAGIFSK+A FFPP  
 Sbjet: 70 KHTGVGLPVVLGCAFSQSVAPLSIIGAQQGSGAMFGALIASGIYVILVAGIFSKIARFFPP 129

Query: 127 IVTGSVITITIGLTLIPVAMGNMGDNKEPSLQSLTSLVLTIGVLLINIFAKGFLKSISI 186  
 IVTGSVIT IGL+L+ VAMGNMGDN KEP+ QS+ LSL+TI ++LL+ F KGF+KSISI  
 Sbjet: 130 IVTGSVITVIGLSLVGVAMGNMGDNVKEPTAQSMMLSLTTIVIIILLVQKFTKGFVKISISI 189

Query: 187 LIGLISGTILAAFMGLVDASVVADAPLVHIPKPFYFGAPRFEFTSILMMCIIATVSMVES 246  
 LIGL++GT+++A MGLVD + V +A +H+P PFYFG P FE TSI+MMCIIATVSMVES  
 Sbjet: 190 LIGLVAGTLVSAMMGLVDTPVVEASWIHVPTPFYFGMPTFEITSIVMMCIIATVSMVES 249

Query: 247 TGVYLALSDITNDKLDKRLRNGYRSEGLAVLLGGLFNTFPYTGFSONVGLVQISGIRTR 306

-2409-

TGVYLALSD+TND+LD KRLRNGYRSEG+AV LGGLFNTFFPYTGFSQNVGLVQISGI+TR  
 Sbjct: 250 TGVYLALSDLTNDQLDEKRLRNGYRSEGIAVFLGGLFNTFFPYTGFSQNVGLVQISGIKTR 309

Query: 307 KPIYFTALFLVILGLLPKFGAMAMIPSPVLGGAMLVLFGMVALQGMKMLNQVDFEHNEH 366  
 +PIY+ A LV++GLLPKF AMAQMIPSPVLGGAMLVLFGMVALQGM+MLN+VDF+ NE+  
 Sbjct: 310 RPIYYAAGILVVIGLLPKFRAMAMIPSPVLGGAMLVLFGMVALQGMQLNRVDFQKNEY 369

Query: 367 NFIIAAVSIAAGVGFNGTNLFIISLPNTLQMFLTNGIVISTLTAVVLNIIILNGLPKK 422  
 NFIIAAVSI+AG+GFNGTNLFI SLP T QMFLTNGIVI+TLT+VVLN++LNG K+  
 Sbjct: 370 NFIIAAVSISAGLGFNGTNLFIASLPETAQMFLTNGIVITATLTSVVLNLVLNGKDKQ 425

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2140

A DNA sequence (GBSx2256) was identified in *Sagalactiae* <SEQ ID 6611> which encodes the amino acid sequence <SEQ ID 6612>. This protein is predicted to be xanthine phosphoribosyltransferase (xpt). Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1921(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA13587 GB:AJ233894 xanthine phosphoribosyltransferase  
 [Streptococcus pneumoniae]  
 Identities = 133/162 (82%), Positives = 144/162 (88%)

Query: 16 GENILKVD SFLTHQVDFELMQEIGKVFADKYKEAGITKVVTTIEASGIAPAVYAAQALGVP 75  
 G+NILKVD SFLTHQVDF LM+EIGKVFA+K+ AGITKVVTTIEASGIAPA++ A+AL VP  
 Sbjct: 1 GDNILKVD SFLTHQVDFSLMREIGKVF AEKFA SAGITKVVTTIEASGIAPALFTA EALNVP 60

Query: 76 MIFAKKAKNITMTEGILTAEVYSFTKQVTSQVSIVSRFLSNDDTVLIIDDFLANGQAAG 135  
 MIFAKKAKNITM EGILTAEVYSFTKQVTS VSI +FLS +D VLIIDDFLANGQAAG  
 Sbjct: 61 MIFAKKAKNITMNEGILTAEVYSFTKQVTS TVSIAGKFLSPEDKVLIIIDDFLANGQAAG 120

Query: 136 LLEIIGQAGAKVAGIGIVIEKSFQDGRDLLEKTGVPVTSLAR 177  
 L++II QAGA V IGIVIEKSFQDGRDLLEK G PV SLAR  
 Sbjct: 121 LIQIIEQAGATVEAIGIVIEKSFQDGRDLLEKAGYPVLSLAR 162

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6613> which encodes the amino acid sequence <SEQ ID 6614>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 156/193 (80%), Positives = 172/193 (88%)

Query: 1 MKLLEERILKDGDLVGENILKVD SFLTHQVDFELMQEIGKVFADKYKEAGITKVVTTIEAS 60  
 M+LLEERIL DG++LGENILKVD+FLTHQVD+ LM+ IGKVFA KY EAGITKVVTTIEAS  
 Sbjct: 1 MQLLEERILTDGNILGENILKVDNFLTHQVDYRLMKAIGKVFAQKYAEAGITKVVTTIEAS 60

-2410-

Query: 61 GIAPAVYAAQALGVPMIFAKKAKNITMTGILTAEVYSFTKQVTSQVSIVSRFLSND DTV 120  
 GIAPAVYAA+A+ VPMIFAKK KNITMTGILTAEVYSFTKQVTS VSI +FLS +D V  
 Sbjct: 61 GIAPAVYAAEAMDVPMIFAKKHKNITMTGILTAEVYSFTKQVTSVSIAGKFLSKEDKV 120

Query: 121 LIIDDFLANGQAAKGLLEIIGQAGAKVAGIGIVIEKSFQDGRDLLEKTGVPVTS LARIKA 180  
 LIIDDFLANGQAAKGL+EIIGQAGA+V G+GIVIEKSFQDGR L+E G+ VTS LARIK  
 Sbjct: 121 LIIDDFLANGQAAKGLLEIIGQAGAQVVGIVIEKSFQDGRRLIEDMGIEVTS LARIKN 180

Query: 181 FENGRVVF AEADA 193  
 FENG + F EADA  
 Sbjct: 181 FENG NLFLEADA 193

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2141

A DNA sequence (GBSx2257) was identified in *S.agalactiae* <SEQ ID 6615> which encodes the amino acid sequence <SEQ ID 6616>. Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2546(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15203 GB:Z99120 similar to GMP reductase [Bacillus subtilis]  
 Identities = 243/321 (75%), Positives = 286/321 (88%), Gaps = 2/321 (0%)

Query: 7 VF DYEDIQLIPNKCIISRSQADTSVKLGNYTFKLPVIPANMQTIIDE EVAETLACEGYF 66  
 VF DYEDIQLIP KCI++SRS+ DTSV+LG +TFKLPV+PANMQTIIDE++A +LA GYF  
 Sbjct: 4 VF DYEDIQLIPAKCIVNSRSECDTSVRLGGHTFKLPVVPANMQTIIDEKLAI SLAENGYF 63

Query: 67 YIMHRFNEEERKPFIKRMHDKGLIASISVGKDY EYDFVTS LKED--APEFTTIDIAHGH 124  
 Y+MHRF E R FIK M+ +GL +SISVGKDY EY+ FV L E+ PE++TIDIAHGH  
 Sbjct: 64 YVMHRFEPETRIDFIKDMNARGLFSSISVGKDEEYEFVRQLAEENLTPEYVTIDIAHGH 123

Query: 125 SNSVIEMIQH I KQELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKVK TGF 184  
 SN+VIEMIQH+K+ LP++FVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITK+KTGF  
 Sbjct: 124 SNAVIEMIQH LKKHLPDSFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKIK TGF 183

Query: 185 GTGGWQLAALRWCSKAARKPIIADGGIRTHGDI AKSIRFGASMVMIGSLFAGHLES PGKL 244  
 GTGGWQLAALRWC+KAA KPIIADGGIRTHGDI AKSIRFGA+MVMIGSLFAGH ESPG+  
 Sbjct: 184 GTGGWQLAALRWCAKAAASKPIIADGGIRTHGDI AKSIRFGATMVMIGSLFAGHEES PGQT 243

Query: 245 VEVEGQQFKEYYGSASEYQKGEHKNVEGKKILLPVKGRLED TLT EMQQDLQSSISYAGGK 304  
 +E +G+ +KEY+GSASE+ KGE KNVEGKK+ + KG ++DTL EM+QDLQSSISYAGG  
 Sbjct: 244 IEKDGLYKEYFGSASEFPKGEKKNVEGKKMHVAHKGSIKDTL IEMEQDLQSSISYAGGT 303

Query: 305 ELDSL RHVDYVIVKNSI WNGD 325  
 +L+++R+VDYVIVKNSI+NGD  
 Sbjct: 304 KLN AIRNV DYVIVKNSI FNGD 324

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6617> which encodes the amino acid sequence <SEQ ID 6618>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-2411-

bacterial cytoplasm --- Certainty=0.2405(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 297/327 (90%), Positives = 311/327 (94%)

Query: 1 MFNDIPVFDYEDIQLIPNKCIISSRSQADTSVKLGNYTFKLPVIPANMQTIIDEEVAETL 60  
 MFNDIPVFDYEDIQLIPNKCII+SRSQADTSV LG Y FKLVPVIPANMQTIIDE +AE L  
 10 Sbjct: 8 MFNDIPVFDYEDIQLIPNKCII+SRSQADTSVTLGKYQFKLPVIPANMQTIIDETIAEQL 67

Query: 61 ACEGYFYIMHRFNEEBERKPFIKRMHDKGLIASISVGKDYEDFVTSLKEDAPEFITIDI 120  
 A EGYFYIMHRF+E+ RKPFIKRMH++GLIASISVGK EY+FVTSLKEDAPEFITIDI  
 15 Sbjct: 68 AKEGYFYIMHRFDEDSRKPFIKRMEQGLIASISVGKACEYEFVTSLKEDAPEFITIDI 127

Query: 121 AHGHSNSVIEIMIQHIKQELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 180  
 AHGH+NSVI+MI+HIK ELPETFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV  
 20 Sbjct: 128 AHGHANSVIDMIKHIKTELPEFVIAGNVGTPEAVRELENAGADATKVGIGPGKVCITKV 187

Query: 181 KTGFGTGGWQLAALRWCSKAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHLES 240  
 KTGFGTGGWQLAALRWC+KAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGH ES  
 25 Sbjct: 188 KTGFGTGGWQLAALRWCAARKPIIADGGIRTHGDIAKSIRFGASMVMIGSLFAGHFES 247

Query: 241 PGKLVEVEGQQFKEYYGSASEYQKGEHKNVEGKKILLPVKGRLEDITLTEMQQDLQSSISY 300  
 PGK VEV+G+ FKEYYGSASEYQKGEHKNVEGKKILLP KG L DTLTEMQQDLQSSISY  
 30 Sbjct: 248 PGKTVEVDGETFKEYYGSASEYQKGEHKNVEGKKILLPTKGLSDTLTEMQQDLQSSISY 307

Query: 301 AGGKELDSLRLHVDYVIVKNSIWNNGDSI 327  
 AGGK+LDLRLHVDYVIVKNSIWNNGDSI  
 30 Sbjct: 308 AGGKLDLRLHVDYVIVKNSIWNNGDSI 334

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2142

35 A DNA sequence (GBSx2258) was identified in *S.agalactiae* <SEQ ID 6619> which encodes the amino acid sequence <SEQ ID 6620>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have an uncleavable N-term signal seq

40 INTEGRAL Likelihood = -16.98 Transmembrane 421 - 437 ( 413 - 443)  
 INTEGRAL Likelihood = -8.81 Transmembrane 166 - 182 ( 159 - 186)  
 INTEGRAL Likelihood = -8.55 Transmembrane 220 - 236 ( 208 - 238)  
 INTEGRAL Likelihood = -6.69 Transmembrane 322 - 338 ( 319 - 353)  
 INTEGRAL Likelihood = -5.26 Transmembrane 199 - 215 ( 196 - 218)  
 45 INTEGRAL Likelihood = -4.35 Transmembrane 343 - 359 ( 342 - 361)  
 INTEGRAL Likelihood = -4.09 Transmembrane 291 - 307 ( 287 - 308)  
 INTEGRAL Likelihood = -3.66 Transmembrane 8 - 24 ( 8 - 27)  
 INTEGRAL Likelihood = -3.66 Transmembrane 133 - 149 ( 133 - 151)  
 INTEGRAL Likelihood = -3.19 Transmembrane 254 - 270 ( 253 - 278)  
 50 INTEGRAL Likelihood = -2.50 Transmembrane 53 - 69 ( 53 - 72)  
 INTEGRAL Likelihood = -1.81 Transmembrane 77 - 93 ( 76 - 95)  
 INTEGRAL Likelihood = -1.33 Transmembrane 109 - 125 ( 109 - 125)

----- Final Results -----

55 bacterial membrane --- Certainty=0.7793(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:CAB61253 GB:AJ250422 ORFC [Oenococcus oeni]  
 Identities = 157/447 (35%), Positives = 252/447 (56%), Gaps = 13/447 (2%)

-2412-

Query: 11 AIIITAILGFSGILLETSMNVTFPLLMEFGVNPVAVIQWVTTGNLLAVAVTVPLSAFMK 70  
 AI+ A L F G+LIETSMNVTFP LM++F ++ +QW+TT LL VA T+ ++AF+ K  
 Sbjct: 15 AILGLAGLAFCGVLIETSMNVTFPTLMQQFSISLNKVQWLTAYLLVAATISIAAFIEK 74

5 Query: 71 NLSEKQIFTLANVLFSLGVLIDSFAPNLAILLVGRVLQGVGTGLALPLLFIILTQIPME 130  
 ++IF A +LF+ GV+ + APN ILL+GR++Q + TGLA+PLL I+ QIP +  
 Sbjct: 75 RFIFKKIFFWAGLLFIIGVICSA LAPNFIILLIGRLIQALSTGLAIPLLITEIMQQIPQK 134

10 Query: 131 RRGMMGVAAAMVTLAPAVGPTYGGVISGMLGWMIFMLLAPILIISTFIGLASIPKRQV 190  
 ++G M + + L P++GPTYGGVI+ L W++IF + PI +I+ IGL+ I ++  
 Sbjct: 135 KQGSYMEIVWELLWQPSLGPTYGGVITQDLWRLLIFWVLPGLIAWLIGLSFIEQKSS 194

15 Query: 191 RINDKLNFFAFISLIGIGLATLLLAIEKMSIF-----YLLVAIVSFVIFYL--NKQ 239  
 + FISL + L ++ +A+ I+ +LL+A++ ++F L N +  
 Sbjct: 195 PSKIPFAWKQFISLILALLSITVAVNNAGIYGWTSIKFYGFLLIAVILLIVFIKLSTNSR 254

20 Query: 240 LEFLNLNVFKDKDFSILLYGVLA FQMIPLALSFLLPNLLQLVLHQTSTKAGLFMFPGAIA 299  
 +++++FK +F L Q I L+L+FLLPN QL+L + +G+ + G++  
 Sbjct: 255 QALISISIFKKWEFVCPLLIYFLIQFIQLSLTFLLPNYAQLILKKGVMISGIMLLCGSLI 314

25 Query: 300 VVFLSPFAGYLLDKIGAFKPIMIGISLSLIGLIGTAIFIPAKSVVLLAFDILTKIGMGI 359  
 L P G +LD P++IG + I IF SV ++ A ++ IG  
 Sbjct: 315 SAILQPLTGRMLDSFSVKIPLVIGAFFLITSTISFTIFQRYLSVFLIAALYVIYMIGFSF 374

30 Query: 360 GASNMVTTALTCLKPAQSADGNSILNTLQQFAGAFATAVASQIFTIGQVAIPKNGAIIGS 419  
 +N +T AL KL +DGN++ NTLQQ+AG+ T+VAS + G K GS  
 Sbjct: 375 VFNNSLTYALQKLPLKLISDGNVFNLTQQYAGSLGTSVASALLANGIGTDGKQSNYTG 434

30 Query: 420 Q--FAVLFVIVVVILAIVGLTYLRKRK 444  
 + F + F+ +++ ++ +K K  
 Sbjct: 435 RHIFILNFISCAIVVILIFSIIQRKKNK 461

There is also homology to SEQ ID 46.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2143

A DNA sequence (GBSx2259) was identified in *Sagalactiae* <SEQ ID 6621> which encodes the amino acid sequence <SEQ ID 6622>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2151(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6595> which encodes the amino acid sequence <SEQ ID 6596>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-2413-

Identities = 74/214 (34%), Positives = 112/214 (51%), Gaps = 5/214 (2%)

Query: 13 NESENNFFITLKYTFNYLFSIQIIT---DISTLNHADFDGSFAFHDIETSIPLVIDSNY 69  
 N+ E F L +F++LF + I+T +I + + F G F+FH+ + +P L ++  
 5 Sbjct: 15 NQLEETFIRELSHHFSLFEVTILTISKANIQSNQLSTFQGIFSFHEHDIDLPTLYFKTSQ 74

Query: 70 LAISQTSNKIEANDIKTFSELSKTMTEFHYMLNFDLFNHLFYRFRHLHNKDGQTIYSNHKP 129  
 ++ + LS+ +T F+ + +LP + RL + +G I NH  
 10 Sbjct: 75 HGQGFLVTESVFDQATAVLSLSQYLTGFYQKFDGHFLQYLPLQARLSDANGNIIVDNHAF 134

Query: 130 EDPFDIYPREEYPIDKWVQNSLIEKKAKELHLLPSASQDYILVQSYKRENDGQVLGY 189  
 F P + I+ W+ L LLPS S D+I +Q Y+ L+N GQLVG  
 15 Sbjct: 135 NGSF--LPTDKEIEDWILAE LR LSNPCKTFLPSGSLDHIYMQHYQALKNPQGQLVGV 192

Query: 190 IEHVHNKIPLEGYLKESGQAIVGWSVDVTSGASI 223  
 ++ V +IKPLL YL+E+GQAIVGWSVDVTSG SI  
 15 Sbjct: 193 LDTVQDIKPLLNQYLEETGQAIVGWSVDVTSGPSI 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 20 vaccines or diagnostics.

**Example 2144**

A DNA sequence (GBSx2260) was identified in *S. agalactiae* <SEQ ID 6623> which encodes the amino  
 acid sequence <SEQ ID 6624>. Analysis of this protein sequence reveals the following:

Possible site: 18

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -12.10 | Transmembrane | 431 - 447 ( 423 - 452) |
| INTEGRAL | Likelihood = -8.92  | Transmembrane | 149 - 165 ( 147 - 174) |
| INTEGRAL | Likelihood = -8.86  | Transmembrane | 404 - 420 ( 402 - 428) |
| INTEGRAL | Likelihood = -7.91  | Transmembrane | 299 - 315 ( 293 - 318) |
| INTEGRAL | Likelihood = -6.42  | Transmembrane | 380 - 396 ( 374 - 398) |
| INTEGRAL | Likelihood = -5.31  | Transmembrane | 350 - 366 ( 347 - 367) |
| INTEGRAL | Likelihood = -4.57  | Transmembrane | 56 - 72 ( 54 - 74)     |
| INTEGRAL | Likelihood = -3.24  | Transmembrane | 172 - 188 ( 171 - 198) |
| INTEGRAL | Likelihood = -1.33  | Transmembrane | 224 - 240 ( 224 - 240) |
| INTEGRAL | Likelihood = -0.59  | Transmembrane | 101 - 117 ( 101 - 117) |

----- Final Results -----

bacterial membrane --- Certainty=0.5840(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF84709 GB:AE004010 potassium uptake protein [Xylella  
 fastidiosa]

Identities = 201/570 (35%), Positives = 319/570 (55%), Gaps = 34/570 (5%)

Query: 1 MAEMQHVNHSSFDKASKAGFII--ALGIVYGDIGTSPLYTMQSLVENQGGISSVTESFIL 58  
 M+ H + ++ G II A+G+V+GDIGTSPLYT++ G++ ++ +L  
 45 Sbjct: 1 MSTSSHSGDCTAVPSNSNGTIILSAIGVVFGDIGTSPLYTLKEAFSPNYGLTPNHDT-VL 59

Query: 59 GSISLIWTLTLITTIKYVLVALKADNHHEGGIFSPLYTLVRKMPW-----LIVPAVI 111  
 G +SLI W + L+ TIKYV V ++ DN EGGI +L L ++ P+ + + +  
 50 Sbjct: 60 GILSLIFWAMMLVVTTIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIYIVGILGIF 119

Query: 112 GGATLLSDGALTTPAVTVTSAVEGLKVVPQLHIFQNSNVIFATLFIILLLLFAIQRFGTG 171  
 G + DG +TPA++V SAVEGL+V F V+ TL +L+LLF QRFGT  
 55 Sbjct: 120 GTSLFFGQGVITPAISVLSAVEGLEVAEPHMKAF-----VVPITLAVLILLFLCQRFGTE 174

Query: 172 VIGKLFQPIWFAFLGISGLLNSFAHPEVFKAINPYGLKLLFSPENHKGIFILGSIF 231  
 +GK FGPI +WF +G+ G+ N PEV AINP +GL F +F+LG++  
 60 Sbjct: 175 RVGKTFGPITLLWFIAIGVVGVIYIAQAPEVLHAINPSWGLH-FFLEHGWHSMFVLGAVV 233

-2414-

Query: 232 LATTGAEALYSIDLGHVGRGNHVSWPVKVAII-LSYCGQAWILANKNAGNELNPFFAS 290  
 LA TG EALY+D+GH G I +W +V + ++ L+Y GQGA +L+N A NPF+ S  
 Sbjct: 234 LAVTGGEALYADMGHFGAKAIRHAWMYVVLPLALNYLGQALVLSNPTAIG--NPFYQS 291

5 Query: 291 IPSQFTMHVVILATLAIIASQALISGSFTLVSEAMRLKIFPQFRSTYPGDN-IGQTYIP 349  
 IP ++ LAT AA+IASQALI+GS++L S+AM+L P+ + + IGQ Y+P  
 Sbjct: 292 IPDWGLYPMIALATAAAVIASQALITGSYSLSSQAMQLGYIPRMNVRHTSQSTIGQIYVP 351

10 Query: 350 VINWFLFAITTSIVLLFKTSAHMEAAAYGLAITITMLMTTILLSFFL-IQKGVKRLVLLM 408  
 +NW L + V+ F S M +AYG+A+T TM++TT+L+ + V R ++ +M  
 Sbjct: 352 TVNWTLLTLVILTIVIGFGDSTSMASAYGVAVTGTMMITTVLMIIYARANPRVPRMLMWM 411

15 Query: 409 MIFFGILEGIFFLASAVKFMHGGYVVVIIAVAIIFIMTIWYKSKIVSRVYKL--LDLKD 466  
 I F ++G FF A+ +KFM G + +++ V I M W +G K++ ++ ++L +  
 Sbjct: 412 AIVFIAVDGAFFYANIIFMDGAWFPLLLGVVIFTFMRTWLRGRKLLHEEMRKDGINLDN 471

20 Query: 467 YIGQLDKLRHDHRYPIYHTNVVYLTNRMEEDMIDKSIMYSILDKRPKKAQVYWFVNIKVT 526  
 ++ L L + P V+LT + ++ ++M+++ + + F+ +K  
 Sbjct: 472 FLPGL-MLAPPVKVP---GTAVFLT--ADSTVVPHALMHNKHNKVLHERNV-FLTVKTL 524

Query: 527 DEPYTA---EYKVDMMGTDFIVKVELYLGF 553  
 PY A K++ + F +V + GF  
 Sbjct: 525 KIPYAANSERLKIEPISNGF-YRVHIRFGF 553

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6625> which encodes the amino acid sequence <SEQ ID 6626>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

30 INTEGRAL Likelihood = -11.78 Transmembrane 428 - 444 ( 421 - 453)  
 INTEGRAL Likelihood = -8.70 Transmembrane 146 - 162 ( 144 - 171)  
 INTEGRAL Likelihood = -7.64 Transmembrane 404 - 420 ( 398 - 426)  
 INTEGRAL Likelihood = -4.88 Transmembrane 296 - 312 ( 294 - 315)  
 INTEGRAL Likelihood = -4.57 Transmembrane 53 - 69 ( 51 - 71)  
 35 INTEGRAL Likelihood = -3.93 Transmembrane 347 - 363 ( 343 - 363)  
 INTEGRAL Likelihood = -2.50 Transmembrane 372 - 388 ( 371 - 388)  
 INTEGRAL Likelihood = -1.33 Transmembrane 169 - 185 ( 169 - 185)  
 INTEGRAL Likelihood = -1.33 Transmembrane 221 - 237 ( 221 - 237)

----- Final Results -----

40 bacterial membrane --- Certainty=0.5713(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:AAF84709 GB:AE004010 potassium uptake protein [Xylella  
 fastidiosa]  
 Identities = 177/467 (37%), Positives = 270/467 (56%), Gaps = 20/467 (4%)

50 Query: 7 TAFDKASKAGFII-ALGIVYGDIGTSPLYTIQSLVENQGGVNQVSESFILGSISLIWTL 65  
 TA S I+ A+G+V+GDIGTSPLYT++ G+ ++ +LG +SLI W +  
 Sbjct: 11 TAVPSNSNGTIIILSAIGVVFGDIGTSPLYTLKEAFSPNYGLTPNHDT-VLGILSLIFWAM 69

55 Query: 66 TLITTIKYVLIALKADNHHEGGIFSLFTLVVRKMSPW-----LIIPAMIGGATLLSDGA 118  
 L+ TIKYV + ++ DN EGGI +L L ++ P+ + I + G + DG  
 Sbjct: 70 MLVVTTIKYVAVIMRVDNDGEGGIMALTALTQRTMPFGSRSIYIVGILGIFGTSLFFGDGV 129

Query: 119 LTPAVTVTSAIEGLKAVPGLSHIYQNQTINVIITTLVILIVLFGIQRFGTGFIGKIFGPVM 178  
 +TPA++V SA+EGL+ + V+ TL +LI+LF QRFGT +GK FGP+  
 Sbjct: 130 ITPAISVLSAVEGLEVAEPHMAF----VVPITLAVLILLFLCQRFGTERVGKTFGPIT 184

60 Query: 179 FIWFSFLGVSGFFNTLGHLEIFKAINPYALHLLFSPENHRGIFILGSIFLATGAEALY 238  
 +WF +GV G +N E+ AINP + LH F +F+LG++ LA TG EALY  
 Sbjct: 185 LLWFIAIGVGVYINIAQAEVLAHAINPSWGLH-FFLEHGWHSMFVLGAVVLAVTGGEALY 243

65 Query: 239 SDLGHVGRGNIYVSWPVKM-CIVLSYCGQAAWILANKHSGIELNPFFASVPSQLRVYLV 297



-2415-

+D+GH G I +W +V + + L+Y GQ A +L+N + NPF+ S+P ++  
 Sbjct: 244 ADMGHFGAKAIRHAWMYVVLPLMLALNYLGQALVLSNPTA--IGNPFYQSIPDWGLYPMI 301

Query: 298 SLATLAIIASQALISGSFTLVSEAMRLKIFPLFRVTPG-ANLGQLYIPVINWILFAVT 356  
 +LAT AA+IASQALI+GS++L S+AM+L P V + + +GQ+Y+P +NW L +  
 Sbjct: 302 ALATAAAVIASQALITGSYSLSQAMQLGYIPRMNVRHTSQSTIGQIYVPTVNWTLTLV 361

Query: 357 SCTVLAFR TSAHMEAAAYGLAITITMLMTTILLKYYLIKKGTRPILAHLVMAF-FALVEFI 415  
 TV+ F S M +AYG+A+T TM++TT+L+ Y P L +MA F V+  
 Sbjct: 362 ILTVIGFGDSTSMASAYGVAVTGTMMITTVLMIIYARANPRVPRMLLWMAIVFIAVDGA 421

Query: 416 FFLASAIKFMHGGYAVVILALAIVFVMIWHAGTRIVFKYVKSLLN 462  
 FF A+ IKFM G + ++L + I M W G +++ + ++ +N  
 Sbjct: 422 FFYANIIKFMGDGAWFPLLLGVVIFTFMRTWLRGRKLLHEEMRKDGIN 468

An alignment of the GAS and GBS proteins is shown below.

Identities = 485/651 (74%), Positives = 575/651 (87%)

Query: 10 SSFDKASKAGFIIALGIVYGDIGTSPLYTMQSLVENQGGISSVTESFILGSISLIITLT 69  
 ++FDKASKAGFIIALGIVYGDIGTSPLYT+QSLVENQGG++ V+ESFILGSISLIITLT  
 Sbjct: 7 TAFDKASKAGFIIALGIVYGDIGTSPLYTIQSLVENQGGVNVQSESFILGSISLIITLT 66

Query: 70 LITTIKYVLVALKADNHHEGGIFSLYTLVRKMPWLIVPAVIGGATLLSDGALTPAVTVT 129  
 LITTIKYVL+ALKADNHHEGGIFSL+TLVRKM+PWL+PA+IGGATLLSDGALTPAVTVT  
 Sbjct: 67 LITTIKYVLIALKADNHHEGGIFSLFTLVRKMSPWLIIPAMIGGATLLSDGALTPAVTVT 126

Query: 130 SAVEGLKVVPSLQHFQNSNVIFATLFIILLFAIQRFGTGVIGKLFPGPIMFIWFAFLG 189  
 SA+EGLK VP L HI+QNQ+NVI TL IL++LF IQRFGTG IGK+FGP+MFIWF+FLG  
 Sbjct: 127 SAIEGLKAVPGLSHIYQNQNTNVIITTLVILIVLFGIQRFGTGFIGKIFGPVPMFIWFSFLG 186

Query: 190 ISGLLNSFAHPEVFKAINPYGLKLLFSPENHKGIFILGSIFLATTGAELYSIDLGHVGR 249  
 +SG N+ H E+FKAINPY L LLFSPENH+GIFILGSIFLATTGAELYSIDLGHVGR  
 Sbjct: 187 VSGFNTLGHLEIFKAINPYALHLLFSPENHRGIFILGSIFLATTGAELYSIDLGHVGR 246

Query: 250 GNIHVSPPFKVAIILSYCGQAWILANKNAGNELNPPFFASIPSQFTMHVILATLAAII 309  
 GNI+VSPPFKV+ I+LSYCGQ AWILANK++G ELNPPFFAS+PSQ +++V LATLAAII  
 Sbjct: 247 GNIYVSPPFKMCIVLSYCGQAAILANKHSGIELNPPFFASVPSQLRVYLVSATLAAII 306

Query: 310 ASQALISGSFTLVSEAMRLKIFPQFRSTYPGDNIGQTYIPVINWFLFAITTSIVLLFKTS 369  
 ASQALISGSFTLVSEAMRLKIFP FR TYPG N+GQ YIPVINW LFA+T+ VL F+TS  
 Sbjct: 307 ASQALISGSFTLVSEAMRLKIFPLFRVTPGANLGQLYIPVINWILFAVTSCTVLAFRS 366

Query: 370 AHMEAAAYGLAITITMLMTTILLSFFLIQGVKRGVLMMIFFGILEGIFFLASAVKFMH 429  
 AHMEAAAYGLAITITMLMTTILL ++LI+KG + L L+M FF ++E IFFLASA+KFMH  
 Sbjct: 367 AHMEAAAYGLAITITMLMTTILLKYYLIKKGTRPILAHLVMAFFALVEFIFFLASAIKFMH 426

Query: 430 GGYVVVIIAVALIIFIMTIWYKGSKIVSRVYKLLDLKDYIGQLDKLRHDHRYPIYHTNVVY 489  
 GGY VVI+A+AI+F+M IW+ G++IV +YVK L+L DY Q+ +LR D + +Y TNVVY  
 Sbjct: 427 GGYAVVILALAIVFVMIWHAGTRIVFKYVKSLLNDYKEQIKQLRDDVCFDLYQTNVVY 486

Query: 490 LITNRMEEDMIDKSIMYSILDKRPKAQVYWFVNIVTDEPYTAEYKVDMMGTDFIVKVEL 549  
 L+NRM++ MID+SI+YSILDKRPK+AQVYWFVN++VTDEPYTA+YKVDMMGTD++V+V L  
 Sbjct: 487 LSNRMQDHMIDRSILYSILDKRPKAQVYWFVNQVTDEPYTAKYKVDMMGTDMVRVNL 546

Query: 550 YLGFMRQTVSRYLRTIVEELLESGRLPKQKTYSVRPDSNVGDFRFRFVLDERFSSSQNL 609  
 YLGF+M QTV RYLRTIV++L+ESGRLPKQ + Y++ P +VGDFRFR++++ER S+++ L  
 Sbjct: 547 YLGFMRPQTVPRYLRTIVQDLMESGRLPKQEQEYTTITPGRDVGDFRFRVLEERVSNAQQL 606

Query: 610 KPGERFVMLMKSSIKHWTATPIRWFLQFSEVTTEVVPLIFTANRGLPIKE 660  
 ERF+M K+SIKH TA+P+RWFLQ+SEVT EVVPLI + LPIKE  
 Sbjct: 607 SNFERFIMQTKASIKHVTASPMRWFLQYSEVTLEVPLILSDVLKLPKE 657

A related GBS gene <SEQ ID 8983> and protein <SEQ ID 8984> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 5.84  
GvH: Signal Score (-7.5): -4.59  
Possible site: 18

```
ALOM program      count: 10 value: -12.10 threshold:  0.0
```

|                           |                     |               |                        |
|---------------------------|---------------------|---------------|------------------------|
| INTEGRAL                  | Likelihood = -12.10 | Transmembrane | 431 - 447 ( 423 - 452) |
| INTEGRAL                  | Likelihood = -8.92  | Transmembrane | 149 - 165 ( 147 - 174) |
| INTEGRAL                  | Likelihood = -8.86  | Transmembrane | 404 - 420 ( 402 - 428) |
| INTEGRAL                  | Likelihood = -7.91  | Transmembrane | 299 - 315 ( 293 - 318) |
| INTEGRAL                  | Likelihood = -6.42  | Transmembrane | 380 - 396 ( 374 - 398) |
| INTEGRAL                  | Likelihood = -5.31  | Transmembrane | 350 - 366 ( 347 - 367) |
| INTEGRAL                  | Likelihood = -4.57  | Transmembrane | 56 - 72 ( 54 - 74)     |
| INTEGRAL                  | Likelihood = -3.24  | Transmembrane | 172 - 188 ( 171 - 198) |
| INTEGRAL                  | Likelihood = -1.33  | Transmembrane | 224 - 240 ( 224 - 240) |
| INTEGRAL                  | Likelihood = -0.59  | Transmembrane | 101 - 117 ( 101 - 117) |
| PERIPHERAL                | Likelihood = 0.85   | 20            |                        |
| modified ALOM score: 2.92 |                     |               |                        |

----- Final Results -----

```

bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

ORF02578 (367 - 1680 of 2607)

GP|9106998|gb|AAF84709.1|AE004010\_6|AE004010(25 - 463 of 634) potassium uptake protein  
{Xylella fastidiosa}

%Match = 17.8

%Identity = 40.4    %Similarity = 63.7

Matches = 177 Mismatches = 150 Conservative Sub.s = 102

180            210            240            270            300            330            360            390  
TSTCLS\*LK\*\*RPGNALIISGLFIDKCCEFNLCYNEFSHFDD\*YYLIGGLAEMQHVNHSSFDKASKAGFIIALGIVYGD  
|:::||  
MSTSSHSGDCTAVPSNSNGTIILSAIGVVFGD  
10                          20                          30

420 450 480 510 540 570 600 612  
 IGTSPLYTMQSLVENQGGISSVTESEFILGSIISLIWTLTLITTIKYVLVALKADNHHEGGIFSLYTLVRKMTP-----W  
 |||||::: :: :: :|| :||| | : : |||| : : || ||| : | :: |  
 IGTSPLYTLKEAFSPNYGLTFPNHDT-VLGILSLIFWAMMLVVTIKYVAVIMRVNDGEGGIMALTALTQRTMPFGSRSIY  
 50 60 70 80 90 100 110

639            669            699            729            759            789            819            849  
LI-VPVIGGATLLSDGALTPAVTVTSAVEGLKVVPSLQHIFQNSNVIFATLFLILLLLFAIQRFGTVIGIKLFGPIMFI  
:: : : | : :: || :|||::| |||||:| : : :| : || :||| :||| :|| ||| :  
IVGILGIFGTSLFFGDGVITPAISVLSAVEGLEV-----AEPHMAKAFVVPITLAVLILLFLCQRFGERVKGKTFGPITLL  
130            140            150            160            170            180

879            909            939            969            999            1029            1059            1089  
WFAFLGISGLNLSFAHPEVFKAINPYYGKLLFSPENHKGIFILGSIFLATTGAELYSDLGHVGRGNIHVSPPFFVKVAI  
|| :| :| :| ||| :||| :|| :: | :||:: || || ||| :| :| | :| :| :| :  
WFIAIGVVGVYNIAQAEVLHAINPSWGLHFFLEHGWH-SMFVLGAVVLAVTGGEALYADMGHFGAKAIRHAWMYVVLPM

200            210            220            230            240            250            260

1116            1146            1176            1206            1236            1266            1296            1326  
I-LSYCGQGAWILANKNAGNELNPFFASIPSQFTMHVVILATLAAIASQALISGSFTLVSEAMRLKIFPQFRSTYPGDN  
: |:| |||| :|: |         |||: ||         :: ||| ||:|||||:|::| |:|:|         |:         :         :  
LALNYLGQCALVLSNPTA--IGNPFYSIPDWGLYPMIALATAAAVIASQALITGSYSLSQQAMQLGYIPRMNVRTSQS

280                  290                  300                  310                  320                  330                  340

1353 1383 1413 1443 1473 1500 1530 1560  
-IQGYTYPVINWFLFAITTSIVLLFKTSAHMEAAAYGLAITITMLMTTILLSFFL-IQKGVKRGLVLLMMIFFGILEGIF  
||| |:| :|| |:| :| :|||:|:| |:|:|:|: : || :: :| || ::| ||  
TIGOIYVPTNVNWTLLTLVILTVIGFGDSTSMASAYGVAVTGTMITTVMIIYARANPRVPRMLMWMMAIVFIAVDGAFF

-2417-

```

          360      370      380      390      400      410      420
1590      1620      1650      1680      1710      1740      1770      1800
5  LASAVKFMHGGYVVVITAVAIIFIMTIWYKSGKIVSRVVKLLDLKDYIGQLDKLRHDHRYPIYHTNVVYLTNRMEEDMID
   |: :||| |: ::: || | | :| |:: ::
YANIIFKMDGAWFPLLLGVVIFTFMRITWLRGRKLLHEEMRKDGINLDNFLPGLMLAPPVKVPGTAVFLTADSTVVPHALM
          440      450      460      470      480      490      500

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2145

A DNA sequence (GBSx2261) was identified in *S.agalactiae* <SEQ ID 6627> which encodes the amino acid sequence <SEQ ID 6628>. This protein is predicted to be serine dehydrogenase. Analysis of this protein sequence reveals the following:

```

15  Possible site: 26
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3261(Affirmative) < succ>
20  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

25  >GP:AAD07424 GB:AE000552 short chain alcohol dehydrogenase
    [Helicobacter pylori 26695]
    Identities = 18/31 (58%), Positives = 25/31 (80%)

    Query: 3  WVASQPEHININRIEIMPVSQTYGQPQPVYRD 33
           W+  QP H+NINRIEIMP+SQT+ P P +++
30  Sbjct: 219 WIYEQLHVNINRIEIMPISQTFAPLPETHKN 249

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6629> which encodes the amino acid sequence <SEQ ID 6630>. Analysis of this protein sequence reveals the following:

```

35  Possible site: 21
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1021(Affirmative) < succ>
40  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 24/33 (72%), Positives = 29/33 (87%)

45  Query: 1  MSWVASQPEHININRIEIMPVSQTYGQPQPVYRD 33
       +SWV  QP H+N+NRIE+MPVSQ+YGPQPV RD
    Sbjct: 20 VSWVIHQPPHVNINRIELMPVSQSYPQPVTRD 52

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2146

A DNA sequence (GBSx2262) was identified in *S.agalactiae* <SEQ ID 6631> which encodes the amino acid sequence <SEQ ID 6632>. Analysis of this protein sequence reveals the following:

-2418-

Possible site: 21  
>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9337> which encodes amino acid sequence <SEQ ID 9338> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10781> which encodes amino acid sequence <SEQ ID 10782> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10951> which encodes amino acid sequence <SEQ ID 10952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA32349 GB:X14130 ORF (AA 1 to 299) [Lactococcus lactis subsp. cremoris]

Identities = 72/215 (33%), Positives = 110/215 (50%), Gaps = 8/215 (3%)

Query: 4 RSKLAAGFLTILMSVATLAACSGKTSNGTN--VVTMKGDTITVSDFYDQVKTSKAAQQSML 61  
+ K+ L + L SG SN T+ V T G +T S FY ++K S + +  
Sbjct: 2 KKKMRLKVLLASTATALLLLSGCQSNQTDQTVATYSGGKVTESSFYKELKQSPTTKTMLA 61  
Query: 62 TLILSRVFDTPQYGDVKVSDKKVSEAYNKTAQGYGNSFSSALSOAGLTPEGYKQOIRTTMLV 121  
+++ R + YG VS K V++AY+ + YG +F + LSQ G + +K+ +RT L  
Sbjct: 62 NMLIYRALNHAYGKSVSTKTVDAYDSYKQYGENFDAFLSQNGFSRSSFKESSLRTNFLS 121  
Query: 122 EYAVKEAAKKELTEANYKEAYKNYPETSVQVIKLDKADKAKSVLKDVKADGADFAKIAK 181  
E A+K+ K+++E+ K A+K Y P+ +VQ I ED AK V+ D+ A G DFA +AK  
Sbjct: 122 EVALKKL--KKVSESQLKAAWKTYQPKVTVQHILTSDEDTAKQVISDLAA-GKDFAMLA 178

Query: 182 E---KTTATDKKVEYKFDSAGTTLPKEVMSAAFKL 213  
T D + F+ TL AA+KL  
Sbjct: 179 TDSIDTATKDNNGGKISFELNNKTLDAFKAAYKL 213

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6633> which encodes the amino acid sequence <SEQ ID 6634>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA25247 GB:M83946 maturation protein [Lactobacillus paracasei]  
Identities = 88/294 (29%), Positives = 146/294 (48%), Gaps = 14/294 (4%)

Query: 7 LIASVVTLASVMALAAACQSTNDNTKVISMKGDTISVSDFYNETKNTEVSQKAMLNLVISR 66  
L+AS T +++ L+ CQS + KV + G ++ S+FY E K + ++ + N++I R  
Sbjct: 10 LLASTAT--ALLLLSGCQSNQADQKVATYSGGKVTESNFYKELKQSPTTKTMLANMLIYR 67  
Query: 67 VFEAQYGDVKVSKKEVEKAYHKTAQYQASFSAAALQSSLTPTFQRQIRSSKLVEYAVKE 126  
YG VS K V AY +QYG +F A L+Q+ + +FK +R++ L E A+K+  
Sbjct: 68 ALNHAYGKSVSTKTVDAYDSYKQYGENFDAFLSQNGFSRSSFKESSLRTNFLSEVALKK 127  
Query: 127 AAKKELTTQEYKAYESYTPMTAVEMITLDNEETAKSVLEELKAEGADFTAIKE---KT 183  
K+++ + K +++Y P + V+ I +E+TAK V+ +L A G DF +AK T  
Sbjct: 128 L--KKVSESQLKAVWKTYQPKVTVQHILTSDEDTAKQVISDL-AAGKDFATLAKTDSIDT 184  
Query: 184 TTPEKKVITYKFDGATNVPTDVVKAASSLNEGGISDVISVLDPTSYQKKFYIVKVTKAE 243

-2419-

T + F+S + AA L G + P + ++K+  
 .Sbjct: 185 ATKDNGGKISFESNNKTLDTFKDAAYKLKNGDYTQT-----PVKVTNGYEVIMINH-P 238  
 Query: 244 KKSDWQEQYKRLKAIILAEKSKDMNFQNKVIANALDKANVKIKDKAFANILAQY 297  
 K + KK L A + A+ S+D + +VI+ L +V IKDK A+ L Y  
 Sbjct: 239 AKGFTTSSKKALTASVYAKWSRDSSIMQRVISQVLKNQHVTIKDKDLADALDSY 292

An alignment of the GAS and GBS proteins is shown below.

Identities = 125/213 (58%), Positives = 168/213 (78%), Gaps = 1/213 (0%)  
 10 Query: 1 MKTRSKLAAGFLTILMSVATLAACSGKTSNGTINVVTMKGDTITVSDFYDQVKTSKAAQQSM 60  
 MK +KL A +TL SV LAAC T++ T V++MKGDTI+VSDFY++ K ++ +Q++M  
 Sbjct: 1 MKNSNKLIA SVVTLASVMALAACQS-TNDNTKVISMKGDTISVSDFYNETKNTEVSQKAM 59  
 15 Query: 61 LTLILSRVFDTQYGDKVSDDKKVSEAYNKTAKGYGNSFSSALSQAGLTPEGYKQQIRTTML 120  
 L L++SRVF+ QYGDKVS K+V +AY+KTA+ YG SFS+AL+Q+ LTPE +K+QIR++ L  
 Sbjct: 60 LNLVISRVFEAQYGDKVSKEVEKAYHKTAEQYGASFSALAQSLLTPETFRQIRSSKL 119  
 20 Query: 121 VEYAVKEAAKKELTEANYKEAYKNYTPETSVQVIKLDKADGADFAKIA 180  
 VEYAVKEAAKKELT YK+AY++YTP +V++I LD E+ AKSVL+++KA+GADF IA  
 Sbjct: 120 VEYAVKEAAKKELTTQEQYKAYESYTPMAVEMITLDNEETAKSVLEELKAEGADFTAIA 179  
 Query: 181 KEKTTATDKKVEYKFDSAGTTLPKEVMSAAFKL 213  
 KEKTT +KKV YKFDS T +P +V+ AA L  
 25 Sbjct: 180 KEKTTTPEKKVYKFDSGATNPVDVKAASSL 212

SEQ ID 10782 (GBS657) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 8-10; MW 62.8kDa) and in Figure 187 (lane 3; MW 63kDa). Purified GBS657-GST is shown in Figure 245, lanes 2 & 3.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 2147

A DNA sequence (GBSx2263) was identified in *S.agalactiae* <SEQ ID 6635> which encodes the amino acid sequence <SEQ ID 6636>. This protein is predicted to be methyltransferase. Analysis of this protein  
 35 sequence reveals the following:

Possible site: 44  
 >>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]  
 Identities = 132/227 (58%), Positives = 169/227 (74%)  
 Query: 1 MVQSYSKNANHNMRPVVKEEIVQYMRQHQQNNGCLAELEAFKQENIPIPHETATYF 60  
 MV++Y +N M RPVVK E+V++MR Q Q G LAE+ FAK+ NIP+IPHET YF  
 50 Sbjct: 1 MVETYKSTSNPMMNRPVKAELVEWMRSSQTQVTGELAEVLNFAKKNIPVIPHETVLYF 60  
 Query: 61 RFLMQTLQPKHILEIGTAIGFSALLMAENAPEAKITTIDRNEEMIALAKENFAKYDHNH 120  
 + L+ L+PK ILEIGTAIGFSAL+MA+ PEA+I TIDRN EMI LAK+N AKYD+ NQ  
 Sbjct: 61 QMLLSLLPKRILEIGTAIGFSALVMAQEVPEAEIVTIDRNPEMIELAKKNLAKYDHRN 120  
 55 Query: 121 ITLLEGDAVDVLQTLDKSYDFVFMDSAKSKYIVFLPQVLKHLVDVGGVVLDLIFQGGDIA 180  
 I L EGDA DVLQ L +D VFMDSAKSKY+ FLP+ L+ L G++++DD+FQ G+I  
 Sbjct: 121 IQLKEGDAADVLQELKGPDLVFMDSAKSKYVEFLPKSLELLSENGLILMDDVDFQAGEIL 180

-2420-

Query: 181 KPIDEVRRGQRTTIYRGLQRLFDSTLQHPDLTATLVPLGDGLLMIRKN 227  
 PI EV+R QR + RGL++LFD +P +++PLGDGLLMI+K+  
 Sbjct: 181 LPIMEVKRNQRALERGLRKLFDDEVFDNPKYMTSVLPLGDGLMIKKH 227

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6637> which encodes the amino acid sequence <SEQ ID 6638>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 10 INTEGRAL Likelihood = -1.38 Transmembrane 153 - 169 ( 152 - 170)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1553(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 15 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA68045 GB:X99710 methyltransferase [Lactococcus lactis]  
 Identities = 134/227 (59%), Positives = 169/227 (74%)

20

Query: 1 MVKSYSKTANHNMRPVVKEELVHYMRTRQKQTTGFLAELEQFARQENIPIIQPEVVAYF 60  
 MV++Y T+N M RPVVK ELV +MR+ Q Q TG LAE+ FA++ NIP+I E V YF  
 Sbjct: 1 MVETYKSTSNPMMNRPVVKAEVEMRSSQTQVTGELAEVLNFAKENNIPVIPHETVLYF 60

25

Query: 61 RFLQLSLQPKHILEIGTAIGFSALLMAENAPDATIVTIDRNREMIDFAKANFAKYDSRQQ 120  
 + LL L+PK ILEIGTAIGFSAL+MA+ P+A IVTIDRN EMI+ AK N AKYD R Q  
 Sbjct: 61 QMLLSLLKPKRILEIGTAIGFSALVMAQEVPEAEIVTIDRNPEMIELAKNLAHYDHRNQ 120

30

Query: 121 IRLLEGDAADILSTLEGNFDFVFMDSAKSKYIVFLPEILRLKVGGVVILDDVFQGGDIT 180  
 I+L EGDAAD+L L+G FD VFMDSAKSKY+ FLP+ L LL G+++DDVFQ G+I  
 Sbjct: 121 IQLKEGDAADVQLKGPFDLVFMDSAKSKYVEFLPKSLELLSENGLILMDDVFQAGEIL 180

35

Query: 181 KPIEDIRRGQRTTIYRGLQSLFDATLTHPNLTSLVPLSDGLLMIRKN 227  
 PI +++R QR + RGL+ LFD +P TS++PL DGLLMI+K+  
 Sbjct: 181 LPIMEVKRNQRALERGLRKLFDDEVFDNPKYMTSVLPLGDGLMIKKH 227

An alignment of the GAS and GBS proteins is shown below.

Identities = 177/235 (75%), Positives = 199/235 (84%)

40

Query: 1 MVQSYSKNANHNMRPVVKEEIVQYMRQHQQKQNGCLAELEAFQENIPIIPHETATYF 60  
 MV+SYSK ANHNMRPVVKEE+V YMR QKQ G LAELE FA+QENIPII E YF  
 Sbjct: 1 MVKSYSKTANHNMRPVVKEELVHYMRTRQKQTTGFLAELEQFARQENIPIIQPEVVAYF 60

45

Query: 61 RFLMQTLQPKHILEIGTAIGFSALLMAENAPEAKITTTIDRNEEMIALAKENFAKYDNHNQ 120  
 RFL+Q+LQPKHILEIGTAIGFSALLMAENAP+A I TIDRN EMI AK NFAKYD+ Q  
 Sbjct: 61 RFLQLSLQPKHILEIGTAIGFSALLMAENAPDATIVTIDRNREMIDFAKANFAKYDSRQQ 120

50

Query: 121 IRLLEGDAVDVLQTLDKSYDFVFMDSAKSKYIVFLPQVLKHLVDVGGVVVILDDVFQGGDIA 180  
 I LLEGDA D+L TL+ ++DFVFMDSAKSKYIVFLP++L+ L VGGVV+LDD+FGGGDI  
 Sbjct: 121 IRLLEGDAADILSTLEGNFDFVFMDSAKSKYIVFLPEILRLKVGGVVILDDVFQGGDIT 180

55

Query: 181 KPIDEVRRGQRTTIYRGLQRLFDSTLQHPDLTATLVPLGDGLLMIRKNADHIVLED 235  
 KPI+++RRGQRTTIYRGLQ LFD+TL HP+LT +LVPL DGLLMIRKN IVL D  
 Sbjct: 181 KPIEDIRRGQRTTIYRGLQSLFDATLTHPNLTSLVPLSDGLLMIRKNQADIVLPD 235

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2421-

**Example 2148**

A DNA sequence (GBSx2264) was identified in *S.agalactiae* <SEQ ID 6639> which encodes the amino acid sequence <SEQ ID 6640>. This protein is predicted to be phosphoglycolate phosphatase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 50
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10   bacterial cytoplasm --- Certainty=0.2193(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8985> which encodes amino acid sequence <SEQ ID 8986> was also identified. This protein appears to be a hydrolase *i.e.* an exposed protein.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
    Identities = 39/117 (33%), Positives = 67/117 (56%), Gaps = 9/117 (7%)

    Query: 98 KEQESRDSKIHLN-PYAKEILEWTKQDIPNFMYTHKGASTHSVLETLQISHYFDEILTG 156
20   KE E+R+ + ++ ++LE Q +F+ +H+ +LE I+ YF E++T
    Sbjct: 25 KENEARELEHPILFEGVSDLLEDILNQGRHFLVSHRNDQVLEILEKTSIAAYFTEVVT 84

    Query: 157 VSGFERKPHPGQGINYLVKRYSLDKSMTTYIGDRPLDLEVAQNAGIKS-----INLR 207
    SGF+RKP+P+ + YL ++Y + + IGDRP+D+E Q AG+ + +NLR
25   Sbjct: 85 SSGFKRKPFPESMLYLREKYQISSGLV--IGDRPIDIEAGQAAGLDTHLFTSIVNLR 139

```

SEQ ID 8986 (GBS240) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 2; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 3; MW 51.5kDa).

30 GBS240-GST was purified as shown in Figure 225, lane 12.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2149**

35 A DNA sequence (GBSx2265) was identified in *S.agalactiae* <SEQ ID 6641> which encodes the amino acid sequence <SEQ ID 6642>. Analysis of this protein sequence reveals the following:

```

    Possible site: 36
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40   bacterial cytoplasm --- Certainty=0.2620(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6643> which encodes the amino acid sequence <SEQ ID 6644>. Analysis of this protein sequence reveals the following:

```

    Possible site: 54
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
50   bacterial cytoplasm --- Certainty=0.2967(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-2422-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 463/599 (77%), Positives = 541/599 (90%)

5  
Query: 1 MSDNRSHEEKYQWDLTTVFATDELWETEVELTQAINAKGFSGHLLDSSQSLEITEV 60  
M+DNRSH+EEKY WDL+T+FATD+ WE EV +L ++ +KGF+GHLLDSS +LL++T+  
Sbjct: 1 MTDNRSHEEKYTWDLSTIFATDKDWEAEVSDLATEVEASKGFAGHLLDSSANLLKVKTK 60

10  
Query: 61 ELDLSRRLEKVVYVYASMKNDQDTTVAKYQEFQAKATALYAKFSETFSFYEPPELLQLESSE 120  
L+L+RR+EKVVYVA MKNDQDTTVAKYQE+QAKA+ LYAKFSE FSFY+PE++ L + D  
Sbjct: 61 YLELARRVEKVVYVYAHMKNDQDTTVAKYQEYQAKASGLYAKFSEVFSFYDPEVMMHLHQED 120

15  
Query: 121 YQSFLLEMPDLQKYDHFKEFIANKPHVLSQNEEELLAGASEIFGAAGETFEILDNADMV 180  
YQ+FL E P+L+ Y+HFF+K+F + HVLSQ EEELLAGA EIF A ETF ILDNAD+V  
Sbjct: 121 YQAFLTETPELVYNHFFDKLFQAREHVLSQAEEELLAGAQEIFNGAETFSILDNADIV 180

20  
Query: 181 FPVVKNAKGEEVELTHGNFISLMESDR+VR+ AY+AMYSTYEQFQHTYAKTLQTNVKSQ 240  
FPVVK N KGE+VELTHGNFISLMES DR+VR+ AY+AMYSTYEQFQHTYAKTLQTNV K Q  
Sbjct: 181 FPVVKNDKGEDVELTHGNFISLMESKDRSVRQAAYEAMYSTYEQFQHTYAKTLQTNVKVQ 240

25  
Query: 241 NFKARVHHYQSARQSALSANFIPEEVYETLIKTNVNHLPLLRHRYMKLRQKVLGLDDLKMY 300  
N+KARVH Y SARQ+A++ANFIPE VY+TL++TVN HPLLRHRY+KLRQ+VLGLDDLKMY  
Sbjct: 241 NYKARVHKYDSARQAAMAANFIPEAVYDTLLETNVNHLPLLRHRYKLRQEVGLDDLKMY 300

30  
Query: 301 DVYTPLSQMDMSFTYDEALKKSEEVLAIFGEAYSERVHRAFTERWIDVHVNGKGRSGAYS 360  
DVYTPLS+ D++ YDEAL+K+E+VLA+FG+ Y++RVHRAFTERWIDVHVNGKGRSGAYS  
Sbjct: 301 DVYTPLSETDLAIGYDEALEKAQKVLAVFGKDYADRVHRAFTERWIDVHVNGKGRSGAYS 360

35  
Query: 361 GGSYDTNMFLLNWQDTLDNLYTLVHETGHSLSHSTFTRENQPYVYGDYSIFLAETSTN 420  
GGSYDTNMF+LLNWQDTLDNLYTLVHETGHSLSHSTFTRE QPYVYGDYSIFLAETSTN  
Sbjct: 361 GGSYDTNMFILLNWQDTLDNLYTLVHETGHSLSHSTFTRETQPYVYGDYSIFLAETSTN 420

40  
Query: 421 ENILTETLLKEVKDDKNRFAILNHYLDGFGKTIFRQTQFAEFEHAIHVADQEGQVLTSEY 480  
ENI+TE LL EV+D+K RFAILNHYLDGF+GT+FRQTQFAEFEHAIH ADQ+G+VLTSEY  
Sbjct: 421 ENIMTEALLNEVQDEKERFAILNHYLDGFRGTFRQTQFAEFEHAIHQADQKGEVLTSEY 480

45  
Query: 481 LNNLYAELNEKYYGLTKEDNHFIQYEWARIPHFYNNYVYFQYATGFAAANYLAERIVNGN 540  
LN LYA+LNEKYYGL+K+DNHFIQYEWARIPHFYNNYV+QYATGFAAA+YLA++IV+G  
Sbjct: 481 LNQLYADLINEKYYGLSKKDNHFIQYEWARIPHFYNNYVYQYATGFAAASYLADKIVHGT 540

Query: 541 PEDKEAYLNLYLKAGNSDYPLNVIKAGVDMTSADYLDAAFRVFEERLVELENLVAKGVH 599  
+D + YL YLK+GNSDYPL VIAKAGVDM DYL+AAF+VF+ERL ELE LV+KG+H  
Sbjct: 541 QDDIDHYLAYLKSGNSDYPLEVIAKAGVDMKGDYLEAAFKVFDERLLEVLVSKGIH 599

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2150**

50 A DNA sequence (GBSx2266) was identified in *S. agalactiae* <SEQ ID 6645> which encodes the amino acid sequence <SEQ ID 6646>. This protein is predicted to be competence protein. Analysis of this protein sequence reveals the following:

Possible site: 22

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

55 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2955 (Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

60 The protein has homology with the following sequences in the GENPEPT database.



-2423-

>GP:AAC23746 GB:AF052209 competence protein [Streptococcus pneumoniae]  
Identities = 127/269 (47%), Positives = 176/269 (65%), Gaps = 8/269 (2%)

5 Query: 1 MLIADKQGNLINLLESHPGKGQYFCPTCCSAVRLKAGRIMRRHFAHISLKNQCFYHENE 60  
M +A+D +G L+N+LE K Y CP C + L+ G +R HFAH SLK+C F+ ENE  
Sbjct: 1 MFVARDARGELVNVLEDKLEKQAYTCAACGGQLHLRQGPSVRTHFAHKSLLKDCDFFENE 60

10 Query: 61 SNEHLQLKAKLYMSLSRENETMLEHHLPEINQIADLFVNETLALE----VQCSRLSEQRL 116  
S EHL K LY L +E + LE+ L E+ QIAD+FVN LALE V C + + L  
Sbjct: 61 SPEHLANKESLYHWLKKETKVQLEYPLSELKQIADVFVNGNLALESSVVVPCPK---KVL 117

15 Query: 117 RERTKAYLQADFQVRWLLGEKLWLKHLRLTNLHKQFLQFSQSIGFHIWELDLRLLEVRLKY 176  
+ER++ Y +QV WLLG+KLWLK RL T L FL FSQ++GF++WELD +VRLKY  
Sbjct: 118 KERSEGYRSQGYQVLLWLLGQKLWLKERLTRLQAGFLYFSQNMGFYVWELDKGKQVRLKY 177

20 Query: 177 LIYEDLRGHVYVLSKTCPL-SGDVLAFLKWPYQSKNLNFYKVKQDRNIRDYVRQQLRYGN 235  
LIY+DLRG ++Y K G +L L+ PY+ + ++ + V +D++I Y+RQQL Y N  
Sbjct: 178 LIYQDLRGKLVHYQIKESYGGQSLLEILRLPYKKQKISHFTVSEDKDICYIRQQLYYQN 237

25 Query: 236 QFWLRKQEKAYLSGQNLITQELMMFFPQI 264  
FW+++Q +AY G+N+LT L ++PQI  
Sbjct: 238 LFWMKQEAAYQKGENILTYGLKEWYPQI 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6647> which encodes the amino acid  
25 sequence <SEQ ID 6648>. Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1034(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 154/312 (49%), Positives = 204/312 (65%), Gaps = 1/312 (0%)

40 Query: 1 MLIADKQGNLINLL-ESHPGKGQYFCPTCCSAVRLKAGRIMRRHFAHISLKNQCFYHEN 59  
+L A D + LI+L+ + K + CP C S VRL+ G I R HFAH+ L +CQF EN  
Sbjct: 4 ILTALDDKNQLISLVTQPISTKPPFRCPACKSPVRLRQGTIRRPFAHVQLAHCQFQAE 63

45 Query: 60 ESNEHLQLKAKLYMSLSRENETMLEHHLPEINQIADLFVNETLALEVQCSRLSEQRLRER 119  
ES EHL LKAKLY SL R +E +LPE+ QIADL+VN+ LALE+QCS L +RL++R  
Sbjct: 64 ESEEHLLTKAKLYTSLVRTEAVCIEKYLPELQIADLWVNDKLALEIQCSPLPVERLKKR 123

50 Query: 120 TKAYLQADFQVRWLLGEKLWLKHLRLTNLHKQFLQFSQSIGFHIWELDLRLLEVRLKYLIY 179  
TKAY + + VRWLLG KLWL LT L KQFL FS S+GFH+WELD +LRLKYLI+  
Sbjct: 124 TKAYQEKGYPVRLWLLGRKLWLNTHLTALQKQFLYFSSSLGFHLWELDAAANLLRLKYLIH 183

55 Query: 180 EDLRGHVYVLSKTCPLSGDVLAFLKWPYQSKNLNFYKVKQDRNIRDYVRQQLRYGNQFWL 239  
EDL G V YL+KT L +++ + PYQ + L Y+ K N+ +++ L + WL  
Sbjct: 184 EDLFGKVSYLTKTISLDHNIMEMFRLPYQQEILYSYQKMTVNLSKRIQRALLARHPKWL 243

60 Query: 240 RQEKAYLSGQNLITQELMMFFPQIQPRVDTFQCITNSLTSFYQNFTNYQKNKNLD 299  
R+QEKAYLSG NLL F+PQ +P + + FCQI +L +Y++F YY+K K+  
Sbjct: 244 RRQEKAYLSGYNLLMLTTDAFYQWRFVQSSSGFCQIKGNLRPYYESFKVYKKEKDKKV 303

Query: 300 QTLYPVVFYDKI 311  
QTL+ P +Y K+  
Sbjct: 304 QTLFSPKYVVKM 315

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
vaccines or diagnostics.

-2424-

**Example 2151**

A DNA sequence (GBSx2267) was identified in *S.agalactiae* <SEQ ID 6649> which encodes the amino acid sequence <SEQ ID 6650>. This protein is predicted to be bicyclomycin resistance protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 25
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -8.33    Transmembrane    78 - 94 ( 75 - 96)
      INTEGRAL    Likelihood = -8.33    Transmembrane   269 - 285 ( 267 - 287)
      INTEGRAL    Likelihood = -7.38    Transmembrane   290 - 306 ( 287 - 314)
10   INTEGRAL    Likelihood = -7.06    Transmembrane   203 - 219 ( 199 - 225)
      INTEGRAL    Likelihood = -6.69    Transmembrane   157 - 173 ( 143 - 184)
      INTEGRAL    Likelihood = -6.42    Transmembrane    53 - 69 ( 44 - 73)
      INTEGRAL    Likelihood = -6.42    Transmembrane   362 - 378 ( 357 - 381)
      INTEGRAL    Likelihood = -3.72    Transmembrane   242 - 258 ( 240 - 261)
15   INTEGRAL    Likelihood = -3.24    Transmembrane   329 - 345 ( 328 - 346)
      INTEGRAL    Likelihood = -1.28    Transmembrane   107 - 123 ( 106 - 123)

    ----- Final Results -----
    bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
20   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

25   >GP:CAA15047 GB:AJ235272 BICYCLOMYCIN RESISTANCE PROTEIN (bcr1)
      [Rickettsia prowazekii]
      Identities = 86/336 (25%), Positives = 159/336 (46%), Gaps = 28/336 (8%)

Query: 73  GKKNVTVLLGLCLILMSGFISFFTSNFSLAMASRLLLIGIGIGLYNSLSISITTDLYEADER 132
      G++ VLLGL + ++S IS F+ N + M +R + G+ + + + S+ D Y+ E
30   Sbjct: 70  GRRPIVLLGLFIYIVSSIISIFSNIEMLMIAIFIQAFGVSVGSVIGQSMARDSYQGAEL 129

Query: 133  ASMIGLRTASLNIGKALTTFIVGLVLA-IGVNYIYLVYLLVIPVFF-FFWKNVPEVENQT 190
      + + + + L AL ++I G ++ + +Y+++ + L + +++ +PE
35   Sbjct: 130  SYVYAILSPWLLFIPALGSYIGGYIIEYLSWHYVFIFPSLAGTILLALYYQILPETNYII 189

Query: 191  HTLKASTTFDT-----KAALLMLITFLVGI---AYIGATVKIPTLLVTKYHYATSFSSNM 242
      ++S F+ K +L L F++G Y G ++ P +L+ + SF +
40   Sbjct: 190  AFSQSSKYFEVFNIIKDKMLWLYAFIIGAFNGIYYGFFIEAPFILIDQMRVLPSPFYGKL 249

Query: 243  LTLIAFSGILVGSVFGKLVK---VFQEKTLILIMILAMGIGNVLFALANNQIIFIVAS--I 297
      LL+F+ I G + G L+K V+ +K + I + G +LEA+ + + FI+ S
45   Sbjct: 250  AFLLSFASIFGGFLGGYLIKRRQVYDKVMSIGFIFSLCGCILFAVDSFILEFILVSNVF 309

Query: 298  LIGASFVGTM-----SSVFFYISKNYAKEHNNFITSALTAGNI-GVILTPLI--LTKLP 349
      I F+ M S+ I+ YA E +T TAG+I G I +I +T
50   Sbjct: 310  AIAMIFMPMMIHMIGHSLIIAITLRYALEDYATVTG---TAGSIFGAIYYVVIASVTYCV 366

Query: 350  SQLHLEPFMTPLITISGLMVINV--FVYLVLMKSNK 383
      S++H E L+ L + +V F Y+ L+ K K
50   Sbjct: 367  SKIHGETISNFSLLCLVLSISSVISFYIYICLLYKKK 402

```

A related GBS gene <SEQ ID 8987> and protein <SEQ ID 8988> were also identified. Analysis of this protein sequence reveals the following:

```

55   Lipop: Possible site: -1    Crend: 7
      McG: Discrim Score:      6.28
      GvH: Signal Score (-7.5): -2.45
      Possible site: 25
      >>> Seems to have a cleavable N-term signal seq.
      ALOM program    count: 10 value: -8.33 threshold: 0.0
60   INTEGRAL    Likelihood = -8.33    Transmembrane    78 - 94 ( 75 - 96)
      INTEGRAL    Likelihood = -8.33    Transmembrane   269 - 285 ( 267 - 287)
      INTEGRAL    Likelihood = -7.38    Transmembrane   290 - 306 ( 287 - 314)

```



-2426-

**Example 2152**

A DNA sequence (GBSx2268) was identified in *S.agalactiae* <SEQ ID 6651> which encodes the amino acid sequence <SEQ ID 6652>. This protein is predicted to be 16S pseudouridylate synthase (rsuA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 52
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2645(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:BAB06992 GB:AP001518 16S pseudouridylate synthase [Bacillus halodurans]
    Identities = 106/234 (45%), Positives = 141/234 (59%), Gaps = 1/234 (0%)

    Query: 1   MRLDKLLGQAGFGSRNQVKKLICSRQVSDGQIVTKDNVIVDSGLQSIFVGKERVCLKES 60
              MR+DK L   GFGSR VKKL+ + V V GQ + + V+ +SI V E V K
    Sbjct: 1   MRIDKFLANMGFGSRKDVKKLLKTGA VRVQQPIKDPSTHVEPESESITVYGEEVEYKPY 60

20   Query: 61 SYLLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPIGRLDRDTEGLLIVTNNGLPGY 120
              Y ++ KP GV+ A D EH+TVIDL+ E+++ P+GRLD+DT GLL++TN+G +
    Sbjct: 61 VYLMNKP KGVICATEDLEHETVIDLLGEEERHYEPSPVGRLDKDTVGLLLITNDGKFNH 120

25   Query: 121 RMLHPKHHVAKTYYYVEVNGFLERDAITFFEEGVVFDGTCCKPAELTIDTANNDKSTARI 180
              ++ PKHHV KTY V G + + + F GVV DDG KPA L I A +S +
    Sbjct: 121 WLMSPKHHPKTYRALVEGHVTEEDVGAFSHGVVLDGYYVTKPATLHILEA-GARSHIEL 179

30   Query: 181 TITEGKFHQVKMFLAYGVKVIYLRISFGDLRLDMNLKPGQYRRLRDSEAAIL 234
              +TEGKFHQVK+MF A G +V+ L RI G+L LD L G+YR L E A+L
    Sbjct: 180 ILTEGKFHQVKRMFQAVGKRVLELERIKIGNLLDP ELARGEYRELTKETALL 233

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6653> which encodes the amino acid sequence <SEQ ID 6654>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 56
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3310(Affirmative) < succ>
40   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 111/194 (57%), Positives = 138/194 (70%)

45   Query: 1   MRLDKLLGQAGFGSRNQVKKLICSRQVSDGQIVTKDNVIVDSGLQSIFVGKERVCLKES 60
              MRLDKLL      GSR+QVKKLI ++ V VD          VD GLQ I V +RV +
    Sbjct: 1   MRLDKLLEGTQVGSRSQVKKLIKAQGVVVDHMPARNGRQNVDPGLQLIEVTGQRVTHPKH 60

50   Query: 61 SYLLLYKPSGVVSAVRDSEHKTVIDLISEKDKVEGLYPIGRLDRDTEGLLIVTNNGLPGY 120
              SY +L KPSGVVSA +D+ + TVID ++E+DK LYP+GRLDRTDEGL+++T+NGPLG+
    Sbjct: 61 SYIILNKPSGVVSAKKDTNYLTVIDQLAEEDKSPDLYPVGRLDRTDEGLVLLTDNGPLGF 120

55   Query: 121 RMLHPKHHVAKTYYYVEVNGFLERDAITFFEEGVVFDGTCCKPAELTIDTANNDKSTARI 180
              RMLHP HHV+KTY V VNG L DA FF G+ F G +C+PA+LTI A+ D+S A +
    Sbjct: 121 RMLHPSHHVSKTYLVTVNGLLAEDASDFFAAGICFPTGEQCQPAQLTILKADTDQSQASL 180

    Query: 181 TITEGKFHQVKMF 194
              TI+EGKFHQVKK F
60   Sbjct: 181 TISEGKFHQVKKCF 194

```

-2427-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2153

A DNA sequence (GBSx2269) was identified in *S.agalactiae* <SEQ ID 6655> which encodes the amino acid sequence <SEQ ID 6656>. Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9745> which encodes amino acid sequence <SEQ ID 9746> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA18872 GB:D90917 hypothetical protein [Synechocystis sp.]  
Identities = 197/318 (61%), Positives = 243/318 (75%)

Query: 22 MGLLVGKWDQWYDTASTGGKFVRTVTQFRHWTKDGSAGPSGDAGFKAESGRYHLYVS 81  
MGLLV+G W DQWYDT STGG+FVR +QFRHW+T DGS GP+G GFKAE+GRYHLYVS  
Sbjct: 1 MGLLVNGIWQDQWYDTESTGGRFVRHDSQFRHWITPDGSPGPTGHGGFKAEGRYHLYVS 60

Query: 82 LACPWASRVLMIRKLKNLESHISISIVNPLMLENGWTFQYKGVIPDMINQSQYLYQIYQ 141  
LACPWA R LI RKLK LE I +S+V+ LM ENGWTF GV+PD + ++YLYQIY  
Sbjct: 61 LACPWAHRTLIFRKLKGLEGMIDVSVVHWMRENGWTFAPGPGVMPDPLFNAEYLYQIYT 120

Query: 142 ASQSDYTGRTVPVPLWDDKFKHTIVNNESEIMRMLNTAFNHITGNTDDYYPDSLQGGIDE 201  
+ + Y+GRVTVP+LWDK+ TIVNNESEI+R+ N+AF+ + + DYYP +L+ QID  
Sbjct: 121 RADAQYSGRVTVPILWDDKQKQITIVNNESEIIRIFNSAFDGLGAKSGDYYPKALRTQIDA 180

Query: 202 MNFIYPKINNGVYKAGFATSQNVYQKEVETLFTALDQLEKHLSDNHLYLVGEQFTEADIR 261  
+N+ IY INNGVYK GFAT+Q Y++ + LF +LD LE L + YL G++ TEAD R  
Sbjct: 181 LNDRIYHTINNGVYKCGFATTQTAYEEAIAPLFESLDWLEGILQGHQYLTGDEITEADWR 240

Query: 262 LFTTLVRFDTVYVYGHFKCNLKHLDYPHLWHYTKRIYNLPGIAETVNFHDHKKHYGSHK 321  
LFTTL+RFD VY GHFKCNL+ + DYP+LW Y + +Y+ PGIAETVNF HIK HYY SH  
Sbjct: 241 LFTTLIRFDVVYVGHFKCNLRRIQDYPNLWRYLRDLYHQPGIAETVNFQHIKGHYESHL 300

Query: 322 TINPTGIIPAGPNLDWTI 339  
INPTGI+P GP LD ++  
Sbjct: 301 NINPTGIVPMGPALDLSL 318

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6656 (GBS655) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 143 (lane 2-4; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2154

A DNA sequence (GBSx2270) was identified in *S.agalactiae* <SEQ ID 6657> which encodes the amino acid sequence <SEQ ID 6658>. Analysis of this protein sequence reveals the following:

Possible site: 43

-2428-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1116(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase  
 [Bacillus subtilis]  
 Identities = 112/243 (46%), Positives = 163/243 (66%), Gaps = 10/243 (4%)

Query: 1 MRVITVKNIDIEGGKIAFTLLLEEKMKAGAQ-T-LGLATGSSPITFYEEIVKS---NLDFSN 55  
 M+++ ++ E K++ +++E+++A LGLATGS+P+ Y++++ +DFS

15 Sbjct: 1 MKILIAEHYEELCKLSAAIIKEQIQAKDAVLGLATGSTPVGlyQLISDYQAGEIDFSK 60

Query: 56 MVSINLDEYVVGIAASNDQSYSYFMHKLHFDAPFKENNL--PNGLAKDLKEEIKRYDAVI 113  
 + + NLDEY G++ S+ QSY++FMH+HLF + +++ P G L+ K Y+ +I

20 Sbjct: 61 VTTFNLDYAGLSPSHPQSYNHFMEHLFQHINMQPDHIHIPQGDNPQLEAACKVYEDLI 120

Query: 114 N-ANPIDFQILGIGRNGHIGFNEPGTTPFDITTHVVDLAPSTIEANSRFFNSIDD-VPKQA 171  
 A ID QILGIG NGHIGFNEPG+ F+ T VV L+ STI+AN+RFF VP+ A

25 Sbjct: 121 RQAGGIDVQILGIGANGHIGFNEPGSDFEDRTRVVKLSESTIQANARFFGGDPVLVPRLA 180

Query: 172 LSMGIGSIMK-SKTIVLVAYGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAA 230  
 +SMGI +IM+ SK IVL+A G EKA+AI M +GP+T D+PASILQKH+ V +I D AA

30 Sbjct: 181 ISMGIKTIMEFSKHIVLLASGEEKADAIQKMAEGPVTTDVPASILQKHNVTVIADYKAA 240

Query: 231 SKL 233  
 KL

30 Sbjct: 241 QKL 243

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6659> which encodes the amino acid sequence <SEQ ID 6660>. Analysis of this protein sequence reveals the following:

35 Possible site: 43  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.59 Transmembrane 174 - 190 ( 174 - 190)

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:CAB12030 GB:Z99105 similar to glucosamine-6-phosphate isomerase  
 [Bacillus subtilis]  
 Identities = 120/244 (49%), Positives = 162/244 (66%), Gaps = 12/244 (4%)

50 Query: 1 MKIIRVQDQIEGGKIAFTLLKDSL-ARKAKTLGLATGSSPISFYQEMVKS----PLDFSD 55  
 MKI+ + E K++ ++K+ + AK LGLATGS+P+ Y++++ +DFS

Sbjct: 1 MKILIAEHYEELCKLSAAIIKEQIQAKDAVLGLATGSTPVGlyQLISDYQAGEIDFSK 60

Query: 56 LTSINLDEYVGLSVESDQSYDYFMRQNLF---NAKPFKKNYLPNGLATDVEAEAKRYNQI 112  
 +T+ NLDEY GLS QSY++FM ++LF N +P ++P G +EA K Y +

55 Sbjct: 61 VTTFNLDYAGLSPSHPQSYNHFMEHLFQHINMQPDHIHIPQGDNPQLEAACKVYEDL 119

Query: 113 IAHP-IDEQVLGIGRNGHIGFNEPGTSFEEETHVVDLQESTIEANSRFFTSIED-VPKQ 170  
 I + ID Q+LGIG NGHIGFNEPG+ FE+ T VV L ESTI+AN+RFF VP+

60 Sbjct: 120 IRQAGGIDVQILGIGANGHIGFNEPGSDFEDRTRVVKLSESTIQANARFFGGDPVLVPR 179

Query: 171 AISMGIASIMK-SEMIVLLAFGQEKADAIAKGMVFGPITEHLPASILQKHDHVIVIVDEAA 229  
 AISMGI +IM+ S+ IVLLA G+EKADAI+ M GP+T +PASILQKH+HV VI D A

Sbjct: 180 AISMGIKTIMEFSKHIVLLASGEEKADAIQKMAEGPVTTDVPASILQKHNVTVIADYKA 239

-2429-

Query: 230 ASQL 233

A +L

Sbjct: 240 AQL 243

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 163/233 (69%), Positives = 201/233 (85%)

Query: 1 MRVITVKNIDIEGGKIAFTLLLEEKMKAGAQTGLATGSSPITFYEEIVKSNLDFSNMVSIN 60  
 M++I V++ IEGGKIAFTLL++ + GA+TLGLATGSSPI+FY+E+VKS LDFS++ SIN  
 10 Sbjct: 1 MKIIRVQDQIEGGKIAFTLLKDSLAKGAKTLGLATGSSPISFYQEMVKSPLDFSGLTSIN 60

Query: 61 LDEYVGIAASNDQSYSYFMHKLFDKPFKNNLPNGLAKDLKEEIKRYDAVINANPIDF 120  
 LDEYVG++ +DQSY YFM ++LF+AKPFK+N LPNGLA D++ E KRY+ +I +PIDF  
 15 Sbjct: 61 LDEYVGLSVESDQSYDYFMRQNLFNKPFKKNYLPNGLATDVEAEAKRYNQIIAEHPIDF 120

Query: 121 QILGIGRNHGIFNEPGTTPFDITTHVVDLAPSTIEANSRFFNSIDDPKQALSMGIGSIM 180  
 Q+LGIGRNHGIFNEPGT F+ THVVDL STIEANSRFF SI+DVPKQA+SMGI SIM  
 Sbjct: 121 QVLGIGRNHGIFNEPGTSFEEETHVVDLQESTIEANSRFFTSIEDVPKQAISMGIASIM 180

20 Query: 181 KSKTIVLVAYGIEKAEAIASMIKGPITEDMPASILQKHDDVVIIVDEAAASKL 233  
 KS+ IVL+A+G EKA+AI M+ GPITE +PASILQKHD V++IVDEAAAS+L  
 Sbjct: 181 KSEMIVLLAFGQEKADAIGMVFPGPITEHLFASILQKHDHVIVIVDEAAASQL 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 25 vaccines or diagnostics.

### Example 2155

A DNA sequence (GBSx2271) was identified in *S.galactiae* <SEQ ID 6661> which encodes the amino acid sequence <SEQ ID 6662>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 30 >>> Seems to have no N-terminal signal sequence

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.12 | Transmembrane | 169 - 185 ( 161 - 194) |
| INTEGRAL | Likelihood = -6.37 | Transmembrane | 151 - 167 ( 145 - 168) |
| INTEGRAL | Likelihood = -5.15 | Transmembrane | 42 - 58 ( 41 - 62)     |
| INTEGRAL | Likelihood = -1.59 | Transmembrane | 207 - 223 ( 207 - 224) |
| INTEGRAL | Likelihood = -1.12 | Transmembrane | 24 - 40 ( 23 - 40)     |

35 ----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.4248(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF13747 GB:AF117351 unknown [Zymomonas mobilis]  
 Identities = 88/216 (40%), Positives = 123/216 (56%)

45 Query: 9 QQLNILRAGVLGANDGIISVAGVVIGVASATHNLWIIFLSAASAILAGAFSMAGGEYVSV 68  
 +Q+ LRA VLGANDGI+S + ++IGVASA + I L+ S ++AGA SMA GEYVSV  
 Sbjct: 17 RQMGWLRASVLGANDGILSTSSLMIGVASAHGSSGNILLAGMSGLIAGALSMAAGEYVSV 76

50 Query: 69 STQKDTEQAAVAREEKLLENPELAKKSLVDIYLAKGESHEHAQWLVDKAFSKNAIEHLV 128  
 S+Q D EQA VARE L+ NP K L +IY+ +G E A + ++ + NA+E +  
 Sbjct: 77 SSQHDMEQADVAREHAELKANPHAELAEIYVERGLDRELALQVAEQLMAHNALEAHL 136

55 Query: 129 EEKYGIEFGEYTSPWHAAISSFIAFAIGSIFPTITILLPFSVRIVGTVIIVIVSLLSTG 188  
 ++ G+ P AA++S I+F+ G+I P +T L P + + +I I+ L G  
 Sbjct: 137 RDELGLTDSLIRPVQAALASAIISFSGGAIVPFLTALFSPPEIINITISLISILCLAVLG 196

Query: 189 YVSAKLGQAPTVPAMRRNVMIIGLTLATYVIGQLF 224  
 V A LG A A R G L M+ T IG F  
 60 Sbjct: 197 MVGAHLGGANVPKALRVTFGCALAMIGTAAIGSFF 232

-2430-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2156

- 5 A DNA sequence (GBSx2272) was identified in *S.agalactiae* <SEQ ID 6663> which encodes the amino acid sequence <SEQ ID 6664>. This protein is predicted to be S-adenosylmethionine tRNA ribosyltransferase (queA). Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3438(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14732 GB:Z99118 S-adenosylmethionine tRNA ribosyltransferase  
[Bacillus subtilis]

Identities = 228/341 (66%), Positives = 279/341 (80%)

Query: 1 MNTNDFDFYLPEELIAQTPLEKRDASKLLVIDHKNKMTDTSFHDHILDELKPGDALVMNN 60  
M + FDF LPE LIAQ PLE+RDAS+L+V+D +TDS F HI+ GD LV+NN  
Sbjct: 1 MKVDLDFDFELPERLIAQVPLEQRDASRLMVLDKHTGELTDSSEFKHIISFFNEGDCLVLNN 60

Query: 61 TRVLPARLYGEKQDTHGHVELLLLKNTGEGDQWEVLAKPAKRLRVGTVKVSFGDGRLIATVT 120  
TRVLPARL+G K+DT VELLLLK GD+WE LAKPAKR++ GT V+FGDGRL A T  
Sbjct: 61 TRVLPARLFGTKEDTGAKVELLLLKQETGDKWETLAKPAKRVKKGTVVTFGDGRLKAICT 120

Query: 121 KELEHGGRIVEFSYDGFLEVLSESLGEMPLPPYIHEKLEDRDRYQTVYAKENGSAAPTA 180  
+ELEHGGR +EF YDGIF EVLESLGEMPLPPYI E+L+D++RYQTVY+KE GSAAAPTA  
Sbjct: 121 EELEHGGRKMEFYDGFYEVLESLGEMPLPPYIKEQLDDKERYQTVYSKEIGSAAPTA 180

Query: 181 GLHFTKELLEKIETKGVKLVLTLHVGLGTFRPVSVNDLDEHEMHSEFYQLSKEAADTLN 240  
GLHFT+E+L++++ KGV++ ++TLHVGLGTFRPVS D ++EH MH+EFYQ+S+E A LN  
Sbjct: 181 GLHFTTEILQQLKDKGVQIEFITLHVGLGTFRPVSADEVEEHNMAEFYQMSEETAALN 240

Query: 241 AVKESGGRIVAVGTTSTRTLETIGSKFNGELKADSGWTNIFIKPGYQFKVVDAFSTNFHL 300  
V+E+GGRI++VGTTSTRTLETI + +G+ KA SGWT+IFI PGY+FK +D TNFHL  
Sbjct: 241 KVRENGGRIISVGTTSTRTLETIAGEHDGQFKASSGWTNIFIPGYEFKAIDGMITNFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLEAYNHAVEERYRFFSFGDAMFV 341  
PKS+L+MLVSA AGR+ +L AYNHAVEE YRFFSFGDAM +  
Sbjct: 301 PKSSLIMLVSALAGRENILRAYNHAVEEYRFFSFGDAMLI 341

- 45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6665> which encodes the amino acid sequence <SEQ ID 6666>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3864(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 297/341 (87%), Positives = 322/341 (94%)



-2431-

Query: 1 MNTNDFDFYLPEELIAQTPLEKRDASKLLVIDHKNKMTDSHFDHILDELKPGDALVMNN 60  
 MNIN+FDL LPEELIAQTPLEKRD+SKLL+IDH+ KTM DSHFDHI+D+L PGDALVMNN  
 Sbjct: 1 MNTNDFDFELPEELIAQTPLEKRDSSKLLIIDHRQKTMVDSHFDHIIDQLNPGDALVMNN 60

5 Query: 61 TRVLPARLYGEKQDTHGHVELLLKNTQEGDQWEVLAKPAKRLRVGTVKVSFGDGRLIATVT 120  
 TRVLPARLYGEK DTHGHVELLLKNT+GDQWEVLAKPAKRL+VG++V+FGDGRL AT+  
 Sbjct: 61 TRVLPARLYGEKPDTHGHVELLLKNTQGDQWEVLAKPAKRLKVGSOVNFQDGRLKATII 120

10 Query: 121 KELEHGGRIVEFSYDGI FLEVLESLGEMPLPPYIHEKLEDRDRYQTVYAKENGSAAPTA 180  
 ELEHGGRIVEFSYDGI FLEVLESLGEMPLPPYIHEKLED +RYQTVYAKENGSAAPTA  
 Sbjct: 121 DELEHGGRIVEFSYDGI FLEVLESLGEMPLPPYIHEKLEDAERYQTVYAKENGSAAPTA 180

15 Query: 181 GLHFTKELLEKIETKGVKLVYLT LHVGLGTFRPVSDNLDEHEMHSEFYQLSKEAADTLN 240  
 GLHFT +LL+KIE KGV LVYLT LHVGLGTFRPVSDNLDEH+MHSEFY LS+FAA TL  
 Sbjct: 181 GLHFTTDLKKIEAKGVHLVYLT LHVGLGTFRPVSDNLDEHDMHSEFYSLSEEAQT LR 240

20 Query: 241 AVKESGGRIVAVGTTSIRTLETIGSKFNGELKADSGWTNIFIKPGYQFKVVDADFSTNFHL 300  
 VK++GGR+VAVGTTSIRTLETIG KF G+++ADSGWTNIFIKPGYQFKVVDADFSTNFHL  
 Sbjct: 241 DVKQAGGRVAVGTTSIRTLETIGGKFGDIQADSGWTNIFIKPGYQFKVVDADFSTNFHL 300

Query: 301 PKSTLVMLVSAFAGRDFVLEAYNHAVEERYRFFSFGDAMFV 341  
 PKSTLVMLVSAFAGRDFVLEAY HAV+E+YRFFSFGDAMFV  
 Sbjct: 301 PKSTLVMLVSAFAGRDFVLEAYRHAVDEKYRFFSFGDAMFV 341

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 2157

A DNA sequence (GBSx2273) was identified in *S. agalactiae* <SEQ ID 6667> which encodes the amino acid sequence <SEQ ID 6668>. Analysis of this protein sequence reveals the following:

30 Possible site: 36  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -14.22 Transmembrane 14 - 30 ( 6 - 34)

35 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6689(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6669> which encodes the amino acid sequence <SEQ ID 6670>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2655(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 126/195 (64%), Positives = 155/195 (78%), Gaps = 1/195 (0%)

Query: 160 MEERFDITETDYEYIGEHNHYVAAFSGAMSIDDMQKYSLVYSENTPAYALAERIGGMDSA 219  
 M ERFDITETDYEY EH+ YVA F+GAMSI DMQ+YSLVYSENTPAYALAER+GGM+ A  
 55 Sbjct: 1 MTERFDITETDYEYDQEHAYVAQFNGAMSIPDMQEYSLVYSENTPAYALAERLGGMNKA 60

Query: 220 YSKFGRYQSKGDIKNIQKNGNKVTTDYIIQVLDYLWKHRKKYDSLITYLEEAFPTDYR 279  
 Y F RYG+ G I I +NGNK+TT YY+QVLDYLW+H+ KY ++ Y+ E+FP YY+  
 Sbjct: 61 YQLFDRYGVSGAITTIDRNGNKITTAYYLQVLDYLWQHDKYKDILYYIGESFPDLYYK 120

-2432-

Query: 280 ALIPSDVVVAQKPGYVREALNVGAIVKEEVPYIVAIYTAGLGGSTQEDSEINGVGLYQLE 339  
 +P V V QKPGYVREALNVGAIV EE PY++A+Y++GLGG+TQ E+NG+G QL  
 Sbjct: 121 TYLP-HVKVYQKPGYVREALNVGAIVCEESPYLIALYSSGLGGATQASEEVNGLGYVQLV 179

Query: 340 QLCFVINQWHRVNMN 354  
 QL +VIN+W+R N+N  
 Sbjct: 180 QLPYVINNEWYRGNLN 194

- 10 SEQ ID 6668 (GBS680) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 10-12; MW 64kDa) and in Figure 239 (lane 9; MW 64 kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 164 (lane 15; MW 40kDa) and in Figure 188 (lane 9; MW 40kDa). Purified GBS680-His is shown in Figure 242, lane 8. Purified GBS680-GST is shown in Figure 246, lanes 6 & 7.
- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2158

A DNA sequence (GBSx2274) was identified in *S.agalactiae* <SEQ ID 6671> which encodes the amino acid sequence <SEQ ID 6672>. Analysis of this protein sequence reveals the following:

20 Possible site: 17  
 >>> Seems to have no N-terminal signal sequence

|             |                    |               |                        |
|-------------|--------------------|---------------|------------------------|
| INTEGRAL    | Likelihood = -4.57 | Transmembrane | 8 - 24 ( 4 - 25)       |
| INTEGRAL    | Likelihood = -2.13 | Transmembrane | 66 - 82 ( 65 - 84)     |
| INTEGRAL    | Likelihood = -1.65 | Transmembrane | 107 - 123 ( 107 - 125) |
| 25 INTEGRAL | Likelihood = -0.69 | Transmembrane | 36 - 52 ( 36 - 52)     |
| INTEGRAL    | Likelihood = -0.48 | Transmembrane | 89 - 105 ( 89 - 105)   |

----- Final Results -----

|    |                     |     |                                |         |
|----|---------------------|-----|--------------------------------|---------|
| 30 | bacterial membrane  | --- | Certainty=0.2826 (Affirmative) | < succ> |
|    | bacterial outside   | --- | Certainty=0.0000 (Not Clear)   | < succ> |
|    | bacterial cytoplasm | --- | Certainty=0.0000 (Not Clear)   | < succ> |

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2159

A DNA sequence (GBSx2275) was identified in *S.agalactiae* <SEQ ID 6673> which encodes the amino acid sequence <SEQ ID 6674>. Analysis of this protein sequence reveals the following:

40 Possible site: 59  
 >>> Seems to have an uncleavable N-term signal seq

|             |                    |               |                        |
|-------------|--------------------|---------------|------------------------|
| INTEGRAL    | Likelihood = -9.87 | Transmembrane | 108 - 124 ( 97 - 133)  |
| INTEGRAL    | Likelihood = -9.08 | Transmembrane | 181 - 197 ( 173 - 201) |
| INTEGRAL    | Likelihood = -7.43 | Transmembrane | 220 - 236 ( 216 - 248) |
| 45 INTEGRAL | Likelihood = -6.69 | Transmembrane | 6 - 22 ( 3 - 28)       |
| INTEGRAL    | Likelihood = -3.72 | Transmembrane | 401 - 417 ( 400 - 417) |
| INTEGRAL    | Likelihood = -3.35 | Transmembrane | 279 - 295 ( 278 - 295) |
| INTEGRAL    | Likelihood = -2.87 | Transmembrane | 31 - 47 ( 30 - 50)     |
| INTEGRAL    | Likelihood = -2.87 | Transmembrane | 244 - 260 ( 242 - 264) |
| 50 INTEGRAL | Likelihood = -0.80 | Transmembrane | 62 - 78 ( 62 - 78)     |

----- Final Results -----

-2433-

bacterial membrane --- Certainty=0.4949(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21770 GB:U32694 H. influenzae predicted coding region HI0092  
 [Haemophilus influenzae Rd]  
 Identities = 232/416 (55%), Positives = 314/416 (74%), Gaps = 3/416 (0%)

10 Query: 4 TFTTTGALIGLALAILLIKKVHPAYSILGALVGGLIGGDLVTIVNTMVLGAQGMSS 63  
 T + GAL+ L +AI LI+KKV PAY +++GALVGGLIGG DL V+ M+ GAQG+ ++  
 Sbjct: 3 TVSAIGALVALIVAIFLILKKVSPAYGMLVGALVGGLIGGADLSQTVSLMIGGAQGITT 62

15 Query: 64 ILRILTSGLAGALIKTGSAEKIAESIKKLGQQRATLAIATMIICAVGVFIDIAVIT 123  
 ++RIL +G+LAG LI++G+A I E+I KLG+ RA+ ALA+ATMI+ AVGVF+D+AVIT  
 Sbjct: 63 VMRILAAGVLAVLIESGAANSITETITNKLGETRALLALALATMILTAVGVFVDVAVIT 122

20 Query: 124 VAPIALAIGKKANLSKSSILLAMIGGGKAGNIISPNPNTIAASEAFKVDLTSLMVQNIIP 183  
 V+PIALA+ ++++LSK++ILLAMIGGGKAGNI+SPNPN IAA++ F + LTS+M+ IIP  
 Sbjct: 123 VSPIALALSRRSDLSKAAILLAMIGGGKAGNIMSPNPNAIAAADTFHLPLTSVMMAGIIP 182

25 Query: 184 AIAALVVTIILAKIVSKKNNDISYDSEEQV--GSDLPAFLPAISGPLVVICLLALRPLFG 241  
 A+ L++T LAK + K + ++ D E V +LP+FL A+ PLV I LLALRPLF  
 Sbjct: 183 ALFGLILTYFLAKRLINKGSKVT-DKEVIVLETQNLPSFLTALVAPLVAILLALRPLFD 241

30 Query: 242 ITIDPLIALPLGGLISILATGYLKETVPFVEYGLSKVVGVSILLIGTGLSGIISKASNLQ 301  
 I +DPLIALPLGGLI G L+ + GLSK+ V+I+L+GTG L+GII S L+  
 Sbjct: 242 IKVDPLIALPLGGLIGAFCMGKLRNINSYAINGLSKMTFPVAILMLGTGALAGIIANSGLK 301

35 Query: 302 FDMIHLLLEFLNMPTFILAPLSGIFMGAATASTTSGTTIASQTFAETLIKSGVPAVSGAAM 361  
 +I LE +P++ILAP+SG+ M ATASTT+GT +AS F+ TL++ GV +++GAAM  
 Sbjct: 302 EVLIQGLEHSGLPYILAPISGVMLSLATASTTAGTAVASNVFSSTLLELGVSSLAGAAM 361

Query: 362 IHAGATVLDLSLPHGSFFHATGGAVNMAIKDRMKLISYEALIGLTSTIVAVVYCYFF 417  
 IHAGATV D +PHGSFFHATGG+VNM IK+R+KLI YE+ +GL TIV+ + + F  
 Sbjct: 362 IHAGATVFDHMPHGSFFHATGGSVNMDIKERLKLIPYESAVGLMMTIVSTLIFGVF 417

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6675> which encodes the amino acid sequence <SEQ ID 6676>. Analysis of this protein sequence reveals the following:

40 Possible site: 51  
 >>> Seems to have an uncleavable N-term signal seq

|             |                     |               |                        |
|-------------|---------------------|---------------|------------------------|
| INTEGRAL    | Likelihood = -11.15 | Transmembrane | 240 - 256 ( 236 - 265) |
| INTEGRAL    | Likelihood = -10.88 | Transmembrane | 3 - 19 ( 1 - 32)       |
| INTEGRAL    | Likelihood = -10.14 | Transmembrane | 269 - 285 ( 263 - 289) |
| 45 INTEGRAL | Likelihood = -7.27  | Transmembrane | 107 - 123 ( 102 - 141) |
| INTEGRAL    | Likelihood = -7.17  | Transmembrane | 307 - 323 ( 303 - 330) |
| INTEGRAL    | Likelihood = -6.64  | Transmembrane | 24 - 40 ( 23 - 43)     |
| INTEGRAL    | Likelihood = -5.63  | Transmembrane | 422 - 438 ( 420 - 442) |
| INTEGRAL    | Likelihood = -3.77  | Transmembrane | 124 - 140 ( 124 - 141) |
| 50 INTEGRAL | Likelihood = -3.24  | Transmembrane | 189 - 205 ( 184 - 207) |
| INTEGRAL    | Likelihood = -2.60  | Transmembrane | 65 - 81 ( 65 - 82)     |
| INTEGRAL    | Likelihood = -2.34  | Transmembrane | 393 - 409 ( 393 - 409) |
| INTEGRAL    | Likelihood = -0.11  | Transmembrane | 149 - 165 ( 149 - 166) |

55 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

>GP:BAB07616 GB:AP001520 unknown conserved protein [Bacillus halodurans]  
 Identities = 155/435 (35%), Positives = 248/435 (56%), Gaps = 21/435 (4%)

Query: 7 LGVLVGIVIVIIYLVKVNIIIAAPLATSLVILFNQMDPTTTLLGKEPNQFMGALSTYIL 66

-2434-

LG+++G++++ L + +II AP+A +V LF +D LL + +M +  
 Sbjct: 2 LGIVLGLVILMLVLAIRGWSIIWVAPIAAGVVALFGGLD----LLPAYTDTYMEGFVNFAK 57  
 Query: 67 NYFAIFLLGSLAKLMETSGATTSIADYILKKVGHDSPIYKVLVAIFLISAILTYGGISLF 126  
 +F +F+LG+I KLME +GA S+A I K +G + ++ + L A+LTYGGISLF  
 Sbjct: 58 QWFPVFMGLAIFGKLMDTGAARSVASAITKLIGTK--RAILGVM LGCAVLTYGGISLF 114  
 Query: 127 VVMFAVLPLARSLFKKMDLAWNLIQVPLWLGIATFTMTILPGTPAIQNVIPIQYLDTSLT 186  
 VV+FA+ PLA +LF++ +++ LI + LG TFTMT +PGTP IQN+IP Y T+  
 Sbjct: 115 VVVFAMYPALALFREANISRRLIPGTIALGAFTFTMTAVPGTPQIQNLIPTSYGTNAM 174  
 Query: 187 AAAIPSIVGSIGCVAFGLFYMKYCLAKSMARGETYATYAFDNEIQVTKNLPFLASILP 246  
 AA + ++ ++ G Y+ + K GE + T + E + + + +P+ S LP  
 Sbjct: 175 AAPMMGVIAALIMGIGGYTYLVWREKKLKEAGE-FTTEPKNGEKEEKEKVENPWL SFLP 233  
 Query: 247 LLLLLIIIALTGSLFGNDFFKNIIFIALAVILTASWLFRQFIPNKIAVFNLGASSSIAP 306  
 L+ +I+ T +L D I +AL++ I+ L + I N GA S+  
 Sbjct: 234 LVSVIV---TLNLLQWD-----IVLALISGIVLIMLLNVGKVKGFQSMNQGAGGSVLA 284  
 Query: 307 IFATASAVAFGAVVMIVPGFTFFSDLIILNIPGNPLISLAVLTSSMSAITGSSSGALGIVM 366  
 I T++AV FG+VV VPGF ++L+L I G+PLIS AV + ++ TGS+SG +GI +  
 Sbjct: 285 IINTSAAVGFSGSVVRVAVPGFERLTTELLGIQSPLISQAVAINVLGATGSASGGMGIAL 344  
 Query: 367 ----PNFAQYYLDQGLNPEMIHRVATIASNIFTIVPQSGVFLTFLALTGLNHNKNAFKETF 422  
 + Q ++ G++PE HRVA+IAS +P +G LT LA+TGL+HK ++K+ F  
 Sbjct: 345 EALGDRYMLAMETGMSPEAFHRVASIASGGLDTLPHNGAVLTLLAITGLSHKESYKDIF 404  
 Query: 423 ITVSSTFIAQVIVI 437  
 + V ++ I  
 Sbjct: 405 VVGCVIPIVSVAFAI 419

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/395 (22%), Positives = 167/395 (42%), Gaps = 40/395 (10%)

Query: 9 GALIGLALAILLIKKVHPAYSILILGALVGGIGGDLVTIV----NTMVLGAQG--MMS 62  
 G L+G+ + I L +K+V+ + L + L D T + +GA +++  
 Sbjct: 8 GVLVGVIIVIIYLVVKEVNIIIAAPLATSLVILFNQMDPTTTLGKEPNQFMGALSTYILN 67  
 Query: 63 SILRILTSGLILAGALIKTGSAEKIAESIKKLGQQ---RAITALAIATMIICAVGVFIDI 119  
 L ILA + +G+ IA+ I+KK+G + + A+ + + I+ G+ + +  
 Sbjct: 68 YFAIFLLGSLAKLMETSGATTSIADYILKKVGHDSPIYKVLVAIFLISAILTYGGISLFV 127  
 Query: 120 AVITVAPIALAIGKKANLSKSSILLAMIGGGKAGNII----SPNPNTIAASEAFKVDLTS 175  
 + V P+A ++ KK +L+ + I + + G + +P + + LT+  
 Sbjct: 128 VMFAVLPLARSLFKKMDLAWNLIQVPLWLGIATFTMTILPGTPAIQNVIPIQYLDTSLT 187  
 Query: 176 LMVQNIIPAIAALVVTII-----LAKIVSKKNNDISY--DSEEQVGS-DLPAFLPAISGP 227  
 + +I+ +I + + LAK +++ +Y D+E QV + +LP FL +I  
 Sbjct: 188 AAIPSIVGSIGCVAFGLFYMKYCLAKSMARGETYATYAFDNEIQVTKNLPFLASILPL 247  
 Query: 228 LVVICLLALRPLFG-----ITIDPLIALPLGLISILATGYLKETVPFVEYGLSKVVG 280  
 L++I + LFG I L+A+ L S L ++ + G S +  
 Sbjct: 248 LLLIIIIALTGSLFGNDFFKNIIFIALAVIL--TASWLFRQFIPNKIAVFNLGASSSIA 305  
 Query: 281 ---VSILLIGTGTLSGIIKASNLFQDMIHLLLEFLNMPTFILAPLSGIFMGAATASTTS GT 337  
 + + G + I+ D+I L P LA L+ M A T S++  
 Sbjct: 306 PIFATASAVAFGAVVMIVPGFTFFSDLI--LNIPGNPLISLAVLTS-SMSAITGSSSGAL 362  
 Query: 338 TIASQTFAETLIKSGVPAVSGAAMIHAGATVLDL 372  
 I FA+ + G+ MIH AT+ ++  
 Sbjct: 363 GIVMPNFAQYYLDQGL----NPEMIHRVATIASNI 393

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-2435-

**Example 2160**

A DNA sequence (GBSx2277) was identified in *S.agalactiae* <SEQ ID 6677> which encodes the amino acid sequence <SEQ ID 6678>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.24 Transmembrane 85 - 101 ( 84 - 101)

----- Final Results -----

bacterial membrane --- Certainty=0.2296(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB16041 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 176/377 (46%), Positives = 234/377 (61%), Gaps = 2/377 (0%)

Query: 1 MKVVVAIDSLKGSLSLEAGNAIKESINEVISGADVEVHPLADGGEGTVEALTLGMGGTI 60  
MK+++A DS K SLS+LEA AI+ V GAD P+ADGGEGTV++L G I

Sbjct: 1 MKIIIPAPDSFKESLSALEAAEAIERGFKSVFPFGADYRKLVPVADGGEGTVQSLVDATNGRI 60

Query: 61 ETIPVKGPLGEKVHASYGII PQRLAIIEMAAAGITLATEERNPLHTTTTGVGEMIKD 120  
V GPLGE V A +G++ + A+IEMAAA+G+ L+ ++RNPL TTT G GE+I

Sbjct: 61 IEQVVTGPLGEPVRAFFGMMGDGRTAVIEMAAASGLHLVPVDKRNPLITTTTRGTGELIGA 120

Query: 121 AISKGCRHFIIGIGGSATNDGGAGMLQALGYALLDKDNQEISLGAQGLADLKSISTDKVI 180  
A+ G IIGIGGSATNDGGAGM+QALG LLD EI G L+ L SI +

Sbjct: 121 ALDAGAERLIIGIGGSATNDGGAGMIQALGGRLLDMSGSEIGPGGALSQLASIDVSGLD 180

Query: 181 BELKECDFKIAICDVINPLCGAQQCSSIFGPQKGADEDMITKMDTWLSNYATLATSVSEKA 240  
L+ ++AC+V NPL G +G +++FGPQKGA DM+ +D +S++A +A

Sbjct: 181 SRLRNVKLEVACNVDNPLTGPKGATAVFGPQKGAADMLDVLQNVSHFADMAEKALGST 240

Query: 241 DATIEGTGAAGGLGF AFLAFTNATLEPGIDIILSEINIEKAISEADLVVTGEGRLDGGQTV 300  
EG GAAGGLG++ L + A L+ GIDI+L ++ E + +ADLV+TGEGR+D QTV

Sbjct: 241 FRDTGAGAGAAGGLGWSLLTYLQADLRGIDIVLEAVDFESIVQDADLVITGEGRIDSQTV 300

Query: 301 MGKAPIGVAKLAKKYGKKVVAFSGSVTEADAILCNQHGIDAFFPIVRRRLISLDEAMSKEVA 360  
GK PIGVAK AK Y V+ +GS++ D+ QHGIDA F IV + L++A

Sbjct: 301 HGKTPIGVAKAAKSYDVPVIGIAGSISRDSNAVYQHGDALFSIVPGAVPLEDAFEHAAE 360

Query: 361 YKNMKETATQVFRNLINL 377

Y M+ TA + I L

Sbjct: 361 Y--MERTARDIAASIKL 375

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6679> which encodes the amino acid sequence <SEQ ID 6680>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.27 Transmembrane 360 - 376 ( 360 - 376)

----- Final Results -----

bacterial membrane --- Certainty=0.1107(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA57927 GB:U18997 ORF\_f408 [Escherichia coli]

Identities = 115/345 (33%), Positives = 182/345 (52%), Gaps = 25/345 (7%)

Query: 24 MKILVAIDSFKGSVTSPELNTSVAQALLSVDKQLVIETRAIADGGEGSLVALSQTAVGRW 83

-2436-

MKI++A DS+K S+++ E+ ++ + + + +ADGGEG++ A+ G  
 Sbjct: 28 MKIIVIAPDSYKESLSASEVAQAIEKGFREIFPDQVSVVPVADGGEGTVEAMIAATQGAE 87  
 Query: 84 HQVKTIDLLRRPIKVAY--YRHAKQAFIESASIIIGIDKITSNSVTYAQATSYGLGLAVKD 141  
 L + ++ K AFIE A+ G++ + + TS G G +  
 Sbjct: 88 RHAWVTGPLGEKVNASWGISGDGKTAFIEMAAASGLELVPAEKRDPVLTTSRGTGELILQ 147  
 Query: 142 AIQKGATQIEIMLGGTGTSDGGKGFLSLESLNYDFMT-----GRSYLDTLASPVTLGL 193  
 A++ GAT I I +GG+ T+DGG G +++L G L+TL + + + GL  
 Sbjct: 148 ALBSGATNIIIGIGGSATNDGGAGMVQALGAKLCDANGNEIGFGGSLNTL-NDIDISGL 206  
 Query: 194 T-----DVTNPYHGPQGFPAVFGPQKGGSLSQIEETDQIASNFAKKVFCQTTI 241  
 DVTNP G G + +FGPQKG S + I E D S++A+ + +  
 Sbjct: 207 DPRLKDCVIRVACDVTNPLVGDNGASRIFGPQKGASEAMIVELDNLSHYAEVIKKALHV 266  
 Query: 242 DLQTI PGSGAAGGLGGAIV-LLGGTTLTSGFSRIAELLNLDNSLQSCDLVITGEGCLDTQS 300  
 D++ +PG+GAAGG+G A++ LG L SG + LNL+ + C LVITGEG +D+QS  
 Sbjct: 267 DVKDPVPGAGAAGGMGAALMAFLGAELKSGIEIVTTALNLEHHIDCTLVITGEGRIDSQS 326  
 Query: 301 QSGKVPVVAIARMAKKYQVPTIALCGSVKIETGLAAEDFL-AVFSI 344  
 GKVP+ +A +AKKY P I + GS+ + G+ + + AVFS+  
 Sbjct: 327 IHGKVPIGVANVAKKYHKPVIGIAGSLTDDVGVVHQHGIDAVFSV 371

An alignment of the GAS and GBS proteins is shown below.

Identities = 128/379 (33%), Positives = 194/379 (50%), Gaps = 23/379 (6%)  
 Query: 1 MKVVVAIDSLKGSLSLEAGNAIKESINEVISGADVEVHPLADGGEGTVEALTLGMGGTI 60  
 MK++VAIDS KGS++S E ++ +++ V +E +ADGGEG++ AL+ + G  
 Sbjct: 24 MKILVAIDSFKGSVTSPELNTSVAQALLSVDKQLVIETRAIADGGEGSLVALSQTAVAGR 83  
 Query: 61 ETIPVKGPLGEKVHASYGIIPQRQLAIEMAAAAGITLIATEERNPLHTTTTYGVGEMIKD 120  
 + L + +Y + A IE A+ GI I + T+YG+G +KD  
 Sbjct: 84 HQVKTIDLLRRPIKVAY--YRHAKQAFIESASIIIGIDKITSNSVTYAQATSYGLGLAVKD 141  
 Query: 121 AISKGRHFIIIGIGGSATNDGGAGMLQALGYALLDKDNQEISLGAQGLADLKSISTDKVI 180  
 AI KG I +GG+ T+DGG G L++L Y + G + L ++++ +  
 Sbjct: 142 AIQKGATQIEIMLGGTGTSDGGKGFLSLESLNYDFMT-----GRSYLDTLASPVTL 190  
 Query: 181 EELKECDFKIAICDVTNPLCGAQCSSIFGPQKGADEDMITKMDTWLSNYATLATSVEKA 240  
 L DVTNP G QG +++FGPQKG I + D SN+A +  
 Sbjct: 191 LGLT-----DVTNPYHGPQGFPAVFGPQKGGSLSQIEETDQIASNFAKKVFCQTTID 242  
 Query: 241 DATIEGTGAAGGLGFAFLAFTNATLEPGIDIILSEINIEKAISEADLVVTGEGRLDGQTV 300  
 TI G+GAAGGLG A + TL G I +N++ ++ DLV+TGEGLD Q+  
 Sbjct: 243 LQTI PGSGAAGGLGGA-IVLLGGTTLTSGFSRIAELLNLDNSLQSCDLVITGEGCLDTQS 301  
 Query: 301 MGKAPIGVAKLAKKYGKKVAFSGSVTEADAILCNQHGDIAFFPIVRRRLISLDEAMSKEVA 360  
 GK P+ +A++AKKY +A GSV + L + + A F I ++ ISL+ A+ K  
 Sbjct: 302 SGKVPVAIARMAKKYQVPTIALCGSVKIETGLAAEDFL-AVFSIQQPISLEAAIDKTTT 360  
 Query: 361 YKNMKETATQVFRILINLYN 379  
 N+K A + LI +N  
 Sbjct: 361 LSNIKILAAANMLLIAQFN 379

SEQ ID 6678 (GBS409) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 7; MW 45.4kDa).

GBS409-His was purified as shown in Figure 214, lane 6.

GBS409d was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 3 & 4; MW 35kDa) and in Figure 188 (lane 12; MW 35kDa). Purified protein is shown in Figure 240, lanes 9-10.

-2437-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2161

A DNA sequence (GBSx2278) was identified in *S.agalactiae* <SEQ ID 6681> which encodes the amino acid sequence <SEQ ID 6682>. Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1886(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21771 GB:U32695 conserved hypothetical protein [Haemophilus influenzae Rd]  
Identities = 97/383 (25%), Positives = 175/383 (45%), Gaps = 52/383 (13%)

Query: 1 MKLRKQLAQQIVTSIKDVCQQDINFINTKGIIFASTNPKRVGEFHEIGLKVAQTGQMIEV 60

M+L K A++IV + +N ++ G+I AS N R+ + H + + +++E+

Sbjct: 1 MQLDKYTAKKIVKRAMKIIHHSVNVMDHDGVIIASGNSTRLNQRHTGAVLALRENRVVEI 60

Query: 61 TD---QESYFGTQAGINIPFYNCCELLATIGISGNPNQVGKYALLAQKMTLILKEHE-L 116

Q+ F Q GIN+P +Y + + +GISG P QV +YA L + LI+++ L

Sbjct: 61 DQALAQKWNFEAQPGINLPIHYLGKNIGVVGISGEPTQVKQYAEIVKMTAEELIVEQQALL 120

Query: 117 DYLDGFRKNEASIVLHHLVEGRELDYYYLNQFLNQYHLSEKTDYRLLTFEINSQKQKLLL 176

+ + R+ + +L L+ LN + ++ + +F++N + +L+

Sbjct: 121 EQESWHRRYKEEFILQ-----LLHCNLNWKEMEQA--KFFSFDLNKSRVVVLI 167

Query: 177 S-----QSEMSLLNFFDK-----LDTAIYTFNYPNQYWLSSDHMFDDYYPNI 219

+ +L+N+ ++ LD + + N +LS M

Sbjct: 168 KLLNPALDNLQNLINYLEQSEFAQDVAILSLDQVVVLKWTQNS--TVLSAQM-----KT 219

Query: 220 LSKFECEKGLYKVGIGQKSSLSLLKR---SYETSILALK-ALKGQK--VNLVDDLDLEL 273

L + K YK+ +G +L L ++ S++++ L LK + + + D+ L +

Sbjct: 220 LLPADYSKQDYKIAVGACNLPLFEQLPLSFQSAQSTLSYGLKHHPRKGIYVDFEHLRPV 279

Query: 274 LLTSIDSNIKQYVLNKAIVNL-SENDKIL---LNSYFKHNLSLKECSQELFIHKNTVQYR 329

LL + + + L K L L SE + IL L YF N L +++LF+H NT++YR

Sbjct: 280 LLAGLSHSWQGNELIKPLSPLFSEENAILYKTLQQYFLSNCDLYLTAEKLFVHPNTLRYS 339

Query: 330 LNKIYESTQLNPRNFKDATLLYL 352

LNKI + T L D LYL

Sbjct: 340 LNKIEQITGLFFNKIDDKLTLLYL 362

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2162

A DNA sequence (GBSx2279) was identified in *S.agalactiae* <SEQ ID 6683> which encodes the amino acid sequence <SEQ ID 6684>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0290(Affirmative) < succ>

-2438-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAF89979 GB:AF206272 beta-glucosidase [Streptococcus mutans]  
 Identities = 334/475 (70%), Positives = 392/475 (82%), Gaps = 8/475 (1%)

Query: 4 FPKHFLWGGAVAANQVEGAFRTDGKGLSVQDVLNPGGLGD-----FTAKPTPDNLKLE 56  
 FP++FLWGGGA AANQ EGA+ DGKGLSVQDV P GG+ T KPT DNLKL

10 Sbjct: 6 FPENFLWGGATAANQFEGAYNQDGKGLSVQDVTPKGGVAQSGSSSPLITEKPTEDNLKLV 65

Query: 57 AIDFYHNYKNDIKLFAEMGFKVFRTSIAWSRIFPNNGDSAPNEAGLQFYDNLFDLLELLKYN 116  
 IDFY+ YK DI LFAEMGFKVFR SIAW+RIFPNNGD PNEAGL FYD +FDEL KY+

15 Sbjct: 66 GIDFYNYRKEDIALFAEMGFKVFRSLIAWTRIFPNNGDDLEPNEAGLAFYDKVFDELA KYD 125

Query: 117 IEPLVTLSHYETPLHLAKTYNGWADRRLIAFFEKFAQTVMERYKDKVKYWLTFNEVNSIL 176  
 IEPLVTLSHYETPLHLA+ YNGWA+R LIAF+E++A+TV RYKDKVKYWLTFNEVNS+L

Sbjct: 126 IEPLVTLSHYETPLHLARKYNGWANRELIAFYERYARTVFTRYKDKVKYWLTFNEVNSVL 185

20 Query: 177 HMPFTSGAIMTDKSQLSPQELYQAIHHELVASARVTKLGRSINPNFKIGCMILAMPAYPM 236  
 H PF SG I+TD QLS Q+LYQA+HHELV SA TK+G INP+FKIGCM+LAMPAYPM

Sbjct: 186 HAPFMSGGIIITDPEQLSKQDLYQAVHHELVVSALATKVGHEINPDFKIGCMVLAMPAYPM 245

25 Query: 237 TSDPRDVLAAARQFEQHNLLFSDIHVRGKYPTYIQSYFKNNGIKIKFEEGDEEVLAQNTVD 296  
 T+DP D LA R+FE N LFSD+H RGKYP YI+ YFK+N I IK EGD+E++ +NTVD

Sbjct: 246 TADPLDQLAVREFENQNYLFSDLHARGKYPNYIKRYFKDNNIDIKMGEQDKELMLENTVD 305

Query: 297 FLSFSYYSVVTQAYDFENYQSGQGNILGGLTNPHLTSEWGWQIDPIGLRLVLNQYYERY 356  
 F+SFSYYSV A++ E+Y SG+GN+LGGL+NP+L SEWGWQIDP+GLRLVLN Y+RY

30 Sbjct: 306 FISFSYYSVAHAHNPEDYNSGRGNVLGGLSNPYLQASEWGWQIDPVGLRLVLNDSYDRY 365

Query: 357 QIPLFIVENGLGAKDQLIETLDGDTVEDDYRIDYMNQHLVQVAKAIEDGVEIMGYTSWG 416  
 Q+PLFIVENGLGAKD L++ DG TVEDDYRIDY+ +HL+QV +A++DGV+++GYT+WG

35 Sbjct: 366 QLPLFIVENGLGAKDVLVQGPDP-TVEDDYRIDYLQKHLMQVGEALQDGVDLLGYTTWG 424

Query: 417 CIDCVSMSTAQLSKRYGLIYVDRNDDGTGSLQRYKKKSFQWYQKVIKTNGQSLFE 471  
 ID VS ST +LSKRYG IYV NDDG+GSL RYKKKSF WY+KVI+TNG SL+E

Sbjct: 425 PIDLVSESTVELSKRYGFIYVACNDDGSGSLARYKKKSFQWYKVIKTNGASLYE 479

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5287> which encodes the amino acid sequence <SEQ ID 5288>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0763(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 390/469 (83%), Positives = 423/469 (90%)

Query: 1 MTFVFPKHFLWGGAVAANQVEGAFRTDGKGLSVQDVLNPGGLGDFTAKPTPDNLKLEAIDF 60  
 M +FPK FLWGGAVAANQVEGAF D KGLSVQDVLNPGGLG++T PT DNL LEAIDF

55 Sbjct: 1 MGIFPKDFLWGGAVAANQVEGAFAEADAKGLSVQDVLNPGGLGEWTDSPSTDNLTLEAIDF 60

Query: 61 YHNYKNDIKLFAEMGFKVFRTSIAWSRIFPNNGDSAPNEAGLQFYDNLFDLLELLKYNIEPL 120  
 YH YK DI LFAEMGFKVFRTSIAWSRIFPNNGD PNEAGLQFYD+LFDELL Y IEPL

60 Sbjct: 61 YHRYKEDIALFAEMGFKVFRTSIAWSRIFPNNGDDQPNEAGLQFYDDLFDELLNYGIEPL 120

Query: 121 VTLSHYETPLHLAKTYNGWADRRLIAFFEKFAQTVMERYKDKVKYWLTFNEVNSILHMPF 180  
 VTLSHYETPLHLAK YNGW DRRLI FFE+FAQTVMERYKDKVKYWLTFNEVNSILHMPF

Sbjct: 121 VTLSHYETPLHLAKAYNGWTDRLIGFFERFAQTVMERYKDKVKYWLTFNEVNSILHMPF 180



-2439-

Query: 181 TSGAIMTDKSQLSPQELYQAIHHELVASARVTKLGRSINPNFKIGCMILAMPAYPMTSDP 240  
 TSG IMT+K +LS Q+LYQAIHHELVASA VTKL INP+ K+GCMILAMPAYPMTSDP  
 Sbjct: 181 TSGGIMTEKEKLSLQDLYQAIHHELVASASVTKLAHEINPDVKVGCMLAMPAYPMTSDP 240

5 Query: 241 RDVLAARQFEQHNLLFSDIHVRGKYPTIYQSYFKNNGIKIKFEEGDEEVLAQNTVDFLSF 300  
 RD+LAA FE NLLFSDIHVRGKYP+YI+SYFK NGI+I FE+GD+E+LA++TVDFLSF  
 Sbjct: 241 RDILAAHAFENLNLFFSDIHVRGKYPSYIKSYFKENGIEIVFEDGDKELLAEHTVDFLSF 300

10 Query: 301 SYYSVTQAYDFENYQSGQGNILGGLTNPHLTTSEWGQIDPIGLRLVLNQYERYQIPL 360  
 SYYSVTQA++ E Y SQQGNILGGL+NP+L +SEWGQIDPIGLRLVLNQYY+RYQIPL  
 Sbjct: 301 SYYSVTQAHNPEAYTSGQGNILGGLSNPYLESSEWGQIDPIGLRLVLNQYYDRYQIPL 360

15 Query: 361 FIVENGLGAKDQLIETLDGDTVEDDYRIDYMNQHLVQVAKAIEDGVEIMGYTSWGCIDC 420  
 FIVENGLGAKDQL++T DG TV DDYRIDYM+QHLVQVAKAIEDGVE+MGYTSWGCIDC  
 Sbjct: 361 FIVENGLGAKDQLVQTADGSMTHDDYRIDYMSQHLVQVAKAIEDGVEVMGYTSWGCIDC 420

Query: 421 VSMSTAQLSKRYGLIYVDRNDDGTGSLQRYKKKSFGWYQKVIKTNGQSL 469  
 VSMSTAQLSKRYG IYVDRNDDGTG L RYKKKSF WY++VI+TNG+ L  
 Sbjct: 421 VSMSTAQLSKRYGFIYVDRNDDGTGQLTRYKKKSFDWYRQVIQTNGRYL 469

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2163

A DNA sequence (GBSx2280) was identified in *S.agalactiae* <SEQ ID 6685> which encodes the amino acid sequence <SEQ ID 6686>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.40 | Transmembrane | 247 - 263 ( 241 - 273) |
| INTEGRAL | Likelihood = -8.55  | Transmembrane | 429 - 445 ( 424 - 450) |
| INTEGRAL | Likelihood = -4.88  | Transmembrane | 285 - 301 ( 280 - 303) |
| INTEGRAL | Likelihood = -3.82  | Transmembrane | 207 - 223 ( 205 - 225) |
| INTEGRAL | Likelihood = -3.40  | Transmembrane | 113 - 129 ( 112 - 139) |
| INTEGRAL | Likelihood = -1.97  | Transmembrane | 309 - 325 ( 305 - 328) |
| INTEGRAL | Likelihood = -1.59  | Transmembrane | 395 - 411 ( 395 - 411) |
| INTEGRAL | Likelihood = -1.49  | Transmembrane | 174 - 190 ( 173 - 193) |

----- Final Results -----

bacterial membrane --- Certainty=0.5161(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA84286 GB:Z34526 beta-glucoside permease [Bacillus subtilis]  
 Identities = 225/594 (37%), Positives = 351/594 (58%), Gaps = 11/594 (1%)

Query: 4 YQETAKAILAAVGGEKNIQHVTCHVTRLRLVLDNDEIVNDQVIKTIPNVIGVMRKNDQYQ 63  
 Y + +K IL VGGE+N+Q V HC+TRLR L ++ + ++ +P V+G +Q+Q  
 Sbjct: 3 YDKLSKDILQLVGGEENVQVRVHCMTLRLFNLDNAKADRSQLEQLPGVMGTNISGEQFQ 62

50 Query: 64 IILGNDVNNYNAFLALGHFENTTREFSSQKKSSILEKLIETIAGVITPLIPALLGGGML 123  
 II+GNDV Y A + + + SS +K ++L + + I+GV TP++PA+ G GM+  
 Sbjct: 63 IILGNDVPKVYQAIVRHSNLSDEKSAGSSSQKKNVLSAVFDVISGVFTPLIPAIAGAGMI 122

55 Query: 124 KVIGILLPMLGIASSSSQTVAFINFFGDAAYFMPIMIAYSASRFKVTPLAATVGGIL 183  
 K + L G + SQ + GD A+YF+P+++A SAA +F P +AA + +  
 Sbjct: 123 KGLVALAVTFGWMAEKSQVHVILTAVGDGAFYFLPLLLAMSAARKFGSNPYVAAIAAAI 182

60 Query: 184 LHPAFVTMVAEGKPLSLFGAPVTLASYGSSVIPILIMVFLMQYIERWINKIVPSVMKSFL 243  
 LHP ++ GKP+S G PVT A+Y S+VIPIL+ +++ Y+E+WI++ + +K +  
 Sbjct: 183 LHPDLTALLGAGKPIISFGLPVTAAYSSTVIPILLSIWIASVEKWIDRFTHASLKLIV 242

Query: 244 QPTLIILISGFLALVVVGPLGVIIGKGLSSAMLSIYHVAPWLALSILGAIMPLVMTGMH 303

-2440-

PT +LI L L+ VGPLG I+G+ LSS + ++ A +A+ +L L++MTGMH  
 Sbjct: 243 VPTFTLLIVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMILLAGTFSLIIMTGMH 302

5 Query: 304 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAVKAKQKQTRQVAFAGLSALLA 363  
 +AF PI + +LPAM +N+ Q AS AV ++++ K+ + +A ++AL+

Sbjct: 303 YAFVPIMINNIAQNGHDYLLPAMFLANMGQAGASFAVFLRSRNNKFKSLALTTSITALM- 361

10 Query: 364 GITEPALYGVTLKFKKPLYAAMISGGLVGAYIGLVNIASYTFVVPSSIIGLPQYINPQGGN 423  
 GITEPA+YGV ++ KKP AA+I G GA+ G+ +ASY +V GLP I G

Sbjct: 362 GITEPAMYGVNMRLLKKPFAAALIGGAAGGAFYGMTGVASY--IVGGNAGLPS-IPVFIGP 418

15 Query: 424 NFSNAVIAAIIATIIITFIITWFLGIDEGENEKSSINAQEHTHIRSGLSKKTLYSPMVGN 483  
 F A+I + + LG ++ ++ S Q H S +E ++SP+ G

Sbjct: 419 TFIYAMIGLVIAFAAETAAYLLGFEDVPSDGSQ---QPAVHEGS----REIIHSPIKGE 471

20 Query: 484 VLPLSKVPDETFSKLLGEGLAITPSVGEVYAPFDGEIISLFPKHAIALKDDKGVEVLI 543  
 V LS+V D FS+ ++G+G AI P GEV +P G + ++F TKHAI + D+G E+LI

Sbjct: 472 VKALSEVKDGVFSAGVMGKFAIEPEEGEVVSPVRGVTITFKTKHAIGITSQDGAELI 531

25 Query: 544 HIGIDTVELNGEGFEQLVKVGDFVKRGQLLLRMDIDFISCKGYSLISPVVVTNS 597  
 HIG+DTV+L G+ F +K GD V G L+ D++ I + GY +I+PV+VTN+

Sbjct: 532 HIGLDTVKLEGQWFTAHIKEGDKVAPGDPLVVSFDLEQIKAAGYDVITPVIVTNT 585

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2883> which encodes the amino acid  
 25 sequence <SEQ ID 2884>. Analysis of this protein sequence reveals the following:

Possible site: 20

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.40 | Transmembrane | 246 - 262 ( 240 - 271) |
| INTEGRAL | Likelihood = -6.26  | Transmembrane | 284 - 300 ( 279 - 304) |
| INTEGRAL | Likelihood = -4.14  | Transmembrane | 173 - 189 ( 172 - 194) |
| INTEGRAL | Likelihood = -3.24  | Transmembrane | 112 - 128 ( 111 - 137) |
| INTEGRAL | Likelihood = -2.39  | Transmembrane | 428 - 444 ( 425 - 445) |
| INTEGRAL | Likelihood = -2.13  | Transmembrane | 383 - 399 ( 380 - 401) |
| INTEGRAL | Likelihood = -1.97  | Transmembrane | 308 - 324 ( 304 - 327) |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5161(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

An alignment of the GAS and GBS proteins is shown below.

Identities = 508/619 (82%), Positives = 561/619 (90%), Gaps = 1/619 (0%)

45 Query: 4 YQETAKAILAAVGGKNIQHVTHCVTRLRLVLDNDEIVNDQVIKTIPNVIGVMRKNDQYQ 63  
 YQETAKAILAAVGG+ NIQ VTHCVTRLRLVL NDE V DQ +K I NVIGVMRKN QYQ

Sbjct: 3 YQETAKAILAAVGGKTNIQVRVTHCVTRLRLVLKNDKVKDQQVKAISNVIGVMRKNGQYQ 62

50 Query: 64 IILGNDVNNYYNAFLALGHFENTREFSSQKKSSILEKLIETIAGVITPLIPALLGGGML 123  
 IILGNDVNNYY AFL+LGHF+N + SS+ K SILE+LIETIAGVITPLIPALLGGGML

Sbjct: 63 IILGNDVNNYYQAFSLSLGHFDNQDEHSSKAKGSILERLIETIAGVITPLIPALLGGGML 122

55 Query: 124 KVIGILLPMLGIASSSSQTVAFINFFGDAAYYFMPIMIAYSAAARFKVTPVLAATVGGIL 183  
 KV+GILLPMLG+AS+ SQTVAFINFFGDAAYYFMP+MIAYSAA+RFKVTPLAAT+ GIL

Sbjct: 123 KVVIGILLPMLGLASADSQTVAFINFFGDAAYYFMPVMIAYSAAARFKVTPVLAATIAGIL 182

60 Query: 184 LHPAFVTMVAEGKPLSLFGAPVTLASYGSSVIPILIMVFLMQYIERWINKIVPSVMKSFL 243  
 LHPAFV MVAEGKPL+LFGAPVT ASYGSSVIPIL+MV+LMQYIE+W+N++VPSVMKSFL

Sbjct: 183 LHPAFVAMVAEGKPLTLFGAPVTPASYGSSVIPILMMVYLMQYIEKWNRLVPSVMKSFL 242

65 Query: 244 QPTLIILISGFLALVVVGPLGVIIGKGLSSAMLSIYHVAPWLALSILGAIMPLVVMTGMH 303  
 QPTLIILISGFLALVVVGPLGVIIG+GLS+ ML+IYHVAPWLAL+ILGAIMPLVVMTGMH

Sbjct: 243 QPTLIILISGFLALVVVGPLGVIIGQGLSNTMLAIYHVAPWLALAILGAIMPLVVMTGMH 302

Query: 304 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAVKAKQKQTRQVAFAGLSALLA 363  
 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVA K KQKQTRQVA AAG+SALLA

-2441-

5 Sbjct: 303 WAFAPIFLAASVATPDVLILPAMLASNLAQGAASLAVAFKTKQKQTRQVALAAGISALLA 362  
 Query: 364 GITEPALYGVTLKFKKPLYAAMISGGLVGAYIGLVNIASYTFVVPSTIIGLPQYINPQGGN 423  
 GITEPALYGVTLKFKKPLYAAMISGGLVGA+IG VNIASYTFVVPSTIIGLPQYINP GG  
 10 Sbjct: 363 GITEPALYGVTLKFKKPLYAAMISGGLVGAFITGFVNIASTYTFVVPSTIIGLPQYINPSGGA 422  
 Query: 424 NFSNAVIAAIAITILTFIITWFLGIDEGENEKSSINAQEHTHIRSGLSKKETLYSPMVGN 483  
 NF+NA+IA ATI+L F +TWF+GIDE E+ K A + + ++SGLS K+TLY+PM G  
 Sbjct: 423 NFTNALIAGTATIVLAFSLTWFMGIDE-ESPKQVSVAADMSQVKSGLSTKQTLYAPMTGE 481  
 15 Query: 484 VLPLSKVPDETFSKLLGEGLAITPSVGEVYAPFDGEIISLFPKHAIALKDDKGVEVLI 543  
 +L LS+VPDETFSKLLGEG AI PS GEVYAPFDGE+I+ FPTKHA+ALK+ +GVEVLI  
 Sbjct: 482 MLFLSEVPDETFSKLLGEGFAILPSEGEVYAPFDGEVITFFPTKHAVALKNTRGVEVLI 541  
 20 Query: 544 HIGIDTVELNGEGFEQLVKVGDFVVRGQQLLLRMDIDFISKGYSLISPVVVTNSIDQLEI 603  
 H+GIDTVEL G+GFEQLV VGD VVRGQ LL+MDIDFI+SKGYSLISPVVVTNS +QLEI  
 Sbjct: 542 HVGIDTVELKGQFEQLVSVGDVVVRGQALLKMDIDFITSKGYSLISPVVVTNSAEQLEI 601  
 Query: 604 IVKDAETMVTNEDDLLVIL 622  
 I++D + MVT ED LLVIL  
 Sbjct: 602 IIQDDKMMVTKEEDALLVIL 620

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 2164

A DNA sequence (GBSx2281) was identified in *S.agalactiae* <SEQ ID 6687> which encodes the amino acid sequence <SEQ ID 6688>. Analysis of this protein sequence reveals the following:

30 Possible site: 18  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1148(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB15944 GB:Z99124 transcriptional antiterminator (BglG family)  
 [Bacillus subtilis]  
 Identities = 118/275 (42%), Positives = 183/275 (65%)  
 Query: 1 MIIKRVLNHNAVISVTHQGLDVLMLMGKGIAPKKRIGDRINSDAIEKSFVLKNSDNMNRFT 60  
 M I +V+N+N + V QG +++MG+G+AF+K+ GD ++ IEK F L N D +F  
 Sbjct: 1 MKIAKVINNNVISVNEQGKELVVMGRGLAFQKKSGDDVDEARIEKVFTLDNKNVSEKFK 60  
 45 Query: 61 ELFITVPEEVVACSERIINLGKIKLGKNLDEILYINLTDHIHSAIERHEQGMVIONPLRL 120  
 L +P E + SE II+ K++LGK L++ +Y++LTDHI+ AI+R+++G+ I+N L  
 Sbjct: 61 TLLYDIPIECMEVSEEIHYAKLQLGKKNLDSIYVSLTDHINFQIRNQKGLDIKNALLW 120  
 50 Query: 121 EIQRYPPEYSIGMKALELIKDELGICLTIDESAFIAMHFVNAGLDNPFNEAHKITEIVS 180  
 E +R Y DE++IG +AL ++K++ G+ L DE+ FIA+H VNA L+ IT+++  
 Sbjct: 121 ETKRLYKDEFAIGKEALVMVKNKTGVSLPEDEAGFIALHIVNAELNEEMPNIINITKVMQ 180  
 Query: 181 YIEQVKV KIDFRTELDESSIDYYRFMTHTKLFAQRVLSGMKYEDDADLLLVKKYPREY 240  
 I VK F+ E +E S+ YYRF+TH K FAQR+ +G E D LL VK+KY R Y  
 55 Sbjct: 181 EILSIVKYHFKIEFNESLHYRFTVTHLKFFAQRLFNQTHMESQDDFLDVTVEKYHRA 240  
 Query: 241 KCVKEIGNNMAIQYQYQLNSSELLYLTVHVKRLVK 275  
 +C K+I + +Y+++L S ELLYLT+H++R+VK  
 60 Sbjct: 241 ECTKKIQTYIEREYEHKITSDELLYLTTHIERVVK 275

-2442-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6689> which encodes the amino acid sequence <SEQ ID 6690>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0680(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 220/279 (78%), Positives = 246/279 (87%)

Query: 1 MIIKRVLNHNNAVISVTHQGLDVLLMGKGI AFKKRIGDRINSDAIEKSFVLKNSDNMNRFT 60

M+IKRVLNHNA IS HQGLD+LLMGKGI F K++GD I +AIE SFVLKNSDNMNRFT

Sbjct: 1 MLIKRVLNHNAAISTNHQGLDILLMGKGITFGKKVGDSEIENAIETSFVLKNSDNMNRFT 60

Query: 61 ELFITVPPEEVVACSERIINLGKIKLGKLNDEILYINLTDHIHSAIERHEQGMVQNPLRL 120

ELFITVP+EVVACSERIINLGKIKLGK LDEILYINLTDHIHSAIERHEQGM+I NPLR

Sbjct: 61 ELFITVPQEVVACSERIINLGKIKLGKTLDEILYINLTDHIHSAIERHEQGMLIHNPLRW 120

Query: 121 EIQRYPPEYSIGMKALELIKDELGICLTIDESAIFAMHFVNAGLDNPFNEAHKITEIVS 180

EIQRYPPEYS+G+KALELI+ LG+ L IDE+AFIAMHFVNA LD PF E H++TEIVS

Sbjct: 121 EIQRYPPEYSLGVKALELIERNLGVTLAIDEAAFIAMHFVNASLDTPFKEPHRLTEIVS 180

Query: 181 YIEQKVKIDFRTELDDESSIDYYRFMTHTKLFAQRVLSGMKYEDDDADLLL VVKKYPREY 240

YIEQK+K DF+TELD++SIDYYRFMTH KLFAQRVLS M Y+DDDA+LLL VVK YP+EY

Sbjct: 181 YIEQKIKTDFKTELDDETSIDYYRFMTHIKLFAQRVLSQMSYDDDAE LLL VVKTKYPKEY 240

Query: 241 KCVKEIGNNMAIQYQQLNSSELLYLTVHVKRLVKNLKE 279

+CV +I + +Y Y LNSSELLYLTVHVKRLVKNLKE

Sbjct: 241 RCVLDISEEIKRNYHLSSELLYLTVHVKRLVKNLKE 279

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2165

A DNA sequence (GBSx2282) was identified in *S.agalactiae* <SEQ ID 6691> which encodes the amino acid sequence <SEQ ID 6692>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1104(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9335> which encodes amino acid sequence <SEQ ID 9336> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6693> which encodes the amino acid sequence <SEQ ID 6694>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3314(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-2443-

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/178 (80%), Positives = 161/178 (90%)

```

5   Query: 1  MTLHHDKHHATYVANANAALKHPEIGEDLEALLADVSIQIPEDIRQAVINNGGGHLNHAL 60
      MTLHHDKHHATYVAN NAALKHPEIGE+LE LLADV++IPEDIRQ +INNGGGHLNHAL
      Sbjct: 24 MTLHHDKHHATYVANTINAALKHPEIGENLEELLADVTKIPEDIRQTLINNGGGHLNHAL 83

10  Query: 61  FWELMSPEETQISQELSEDINATFGSFEDFKAAFTAAATGRFGSGWAWLVVNAEGKLEVL 120
      FWEL+SPE+ ++ ++++ I+ FGSF+ FK FTAAATGRFGSGWAWLVVN EG+LE+
      Sbjct: 84 FWELLSPEKQDVTDPDVAQAIDDAFGSFDAFKEQFTAAATGRFGSGWAWLVVNKEGQLEIT 143

      Query: 121 STANQDTPIMEGKKPILGLDVWEHAYYLNRYNVRPNYIKAFFEINWNKVNELYQAAK 178
      STANQDTP I EGKKPIL LDVWEHAYYLNRYNVRPNYIKAFFE I+NW KV+ELYQAAK
15  Sbjct: 144 STANQDTPISEGKKPILALDVWEHAYYLNRYNVRPNYIKAFFEIVNWKKVSELYQAAK 201

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2166

20 A DNA sequence (GBSx2283) was identified in *S.agalactiae* <SEQ ID 6695> which encodes the amino acid sequence <SEQ ID 6696>. Analysis of this protein sequence reveals the following:

Possible site: 39  
>>> Seems to have no N-terminal signal sequence

```

25  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3331(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2167

35 A DNA sequence (GBSx2284) was identified in *S.agalactiae* <SEQ ID 6697> which encodes the amino acid sequence <SEQ ID 6698>. This protein is predicted to be DNA polymerase III delta subunit. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence

```

40  ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0511(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9743> which encodes amino acid sequence <SEQ ID 9744> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6699> which encodes the amino acid sequence <SEQ ID 6700>. Analysis of this protein sequence reveals the following:

50 Possible site: 43

-2444-

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.22 Transmembrane 250 - 266 ( 249 - 266)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1489(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 222/340 (65%), Positives = 282/340 (82%)

Query: 1 MIAIEEIGRITPDNLGLVTVLAGEDLGQYAQMKEKLFQVIGFNKDDLAYSYFDLSEEDYQ 60  
 MIAIE+I +++ +NLGL+T++ G+D+GQY+Q+K +L + I F+KDDLAYSYFD+SE YQ  
 Sbjct: 1 MIAIEKIEKLSKENLGLITLVTGDDIGQYSQKSRIMEQIAFDKDDLAYSYFDMSEAAAYQ 60

15 Query: 61 NAELDLESPLPFLSDYKVVIFDQFQDITTDKKTYLDEQAMKRFEAYLQNPVDTTTRLVICAP 120  
 +AE+DL SLPF ++ KVVIFD DITT+KK++L E+ +K FEAYL+NP++TTRL+I AP  
 Sbjct: 61 DAEMDLVSLPFFAEQKVVIFDHLDDITTNKKSFLKEKDLKAFEAYLENPLETTTRLIIFAP 120

20 Query: 121 GKLDGKRRLVKLLKRDARVLEANTLKESDLKTYFQKYAHQEGLVFEAGVFDELLIKSNYD 180  
 GKLD KRRLVKLLKRDA VLEAN LKE++L+TYFQKY+HQ GL FE+G FD+LL+KSN D  
 Sbjct: 121 GKLDKRRLVKLLKRDALVLEANPLKEAELRTYFQKYSHQLGLGFESGAFDQLLLKSNDD 180

25 Query: 181 FSDTLTNIAFLKSYKTDGHISNDVREAI PKSLQDNIFDLTQDVLLGRIDLARDLVRDLR 240  
 FS + N+AFLK+YK G+IS D+ +AIPKSLQDNIFDLT+ VL G+ID ARDL+ DLR  
 Sbjct: 181 FSQIMKNMAFLKAYKKTGNISLTDIEQAIPKSLQDNIFDLTRLVLGGKIDAARDLIHDLR 240

30 Query: 241 LQGEDEIKLIAIMLGQFRMFLQVKILASKGKSESQIVSEL SHYIGRKINPYQVKFAVRDS 300  
 L GED+IKLIAIMLGQFR+FLQ+ ILA K+E Q+V LS +GR++NPYQVK+A++DS  
 Sbjct: 241 LSGEDDIKLIAIMLGQFRLFLQLTILARDVKNEQQLVISLSDILGRRVNPYQVKYALKDS 300

35 Query: 301 RNLPLAFLKEAIRILIEDYAIKRGTYDKDYLFDLALLKI 340  
 R L LAFL A++ LIETDY IK G Y+K YL D+ALLKI  
 Sbjct: 301 RPLSLAFLTGAVKTLIETDYQIKTGLYEKSYLVDIALLLKI 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2168

40 A DNA sequence (GBSx2285) was identified in *S.agalactiae* <SEQ ID 6701> which encodes the amino acid sequence <SEQ ID 6702>. Analysis of this protein sequence reveals the following:

Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3071(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2169

55 A DNA sequence (GBSx2286) was identified in *S.agalactiae* <SEQ ID 6703> which encodes the amino acid sequence <SEQ ID 6704>. This protein is predicted to be esterase. Analysis of this protein sequence reveals the following:

-2445-

Possible site: 26

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -0.32 Transmembrane 175 - 191 ( 175 - 191)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1128(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

10 The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:AAB17013 GB:L38252 esterase [Acinetobacter lwoffii]

Identities = 63/218 (28%), Positives = 107/218 (48%), Gaps = 3/218 (1%)

15 Query: 105 KVIFYVHGGSYIHQASELQYIFVNKLAKKLDAKVVFPIYPKAPTNYNSDAIPKIKLYQN 164  
++IF++HGG++ + LA + +V+ YP AP + Y +AI I +YQ

Sbjct: 73 QLIFHIHGGAFGLGSLNTHRALMTDLAARTQMQUIHVDYPLAPEHPYPEAIDAIFDVYQA 132

Query: 165 TLASVTSPPKQIILVGESAGGGLALGLADNLVTEHIKQPKKEIILISPWLDIATNNPKIEKV 224

20 L PK II+ G+S G LAL L L + P +IL+SP+LD+ + +  
Sbjct: 133 LLVQGIKPKDIIISGDSGANLALALCLRLKQQPELMPGSLILMSPYLDLTLTSESLRFN 192

Query: 225 QKQDPLLKAWQLQQVAPYWANGKKNFKNPQVSPLYSSQFNKMAPIISFFIGTHDIFYPDNQ 284

25 QK D LL LQ ++ +P+VSPL+ + + P +G+ +I D++  
Sbjct: 193 QKHDALLSIEALQAGIKHYLTDDIQPGDPRVSPLF-DDLDGLPPTLVQVGSKEILLDDSK 251

Query: 285 LLHQKLAKENIKHHYIVGQKMNHVYPVLP--IPEAETA 320

+K + ++K H+ + M H + + PEA+ A

Sbjct: 252 RFREKAEQADVKVHFVFKLYTGMWHNFQMFNAWPFPEAKQA 289

30 There is also homology to SEQ ID 3498.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2170**35 A DNA sequence (GBSx2287) was identified in *S.agalactiae* <SEQ ID 6705> which encodes the amino acid sequence <SEQ ID 6706>. This protein is predicted to be purine nucleotide synthesis repressor. Analysis of this protein sequence reveals the following:

Possible site: 51

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

40 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2970(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

45 The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CAB16124 GB:Z99124 similar to transcriptional regulator (LacI family) [Bacillus subtilis]

Identities = 111/300 (37%), Positives = 175/300 (58%), Gaps = 4/300 (1%)

50 Query: 1 MTSISDIAKKAGVAKSTVSRVINHHHPVSDETRQKVMALITELDYIPNQLARDLSRGKTQ 60

M +I +IA+ A V+ STVSRV+NHHP+VS+E R+ V ++ ELDY PN+ A DL RGKT

Sbjct: 1 MANIKETIARLANVSVSTVSRVLNHHHPYVSEKRLVHQVMKELDYTPNRTAIDLIRGKTH 60

55 Query: 61 KIGVVIPIHTRHPYFTQLINGLLDAAKTTDYQLVMPSPDYNQEELSILKQKMEIDALI 120

+GV++P++ HP F +++NG+ AA +Y ++P++YN ++E+ YL+ L+ + ID LI

Sbjct: 61 TVGVILPYSDHPCFDKIVNGITKAAFQHEYATTLLPTNYPNDIETKYLELLRTKKIDGLI 120

Query: 121 FTSRAISLDIIETIYAKYGRIVVCEKIQEYNHLSSAYLDRYSSFLEAFSDMKLRGLEHLVL 180

TSRA D I Y +YG ++ CE + + + A+ DR +++ E+F +K RG E++

-2446-

Sbjct: 121 ITSRAHWDSILAYQEYGPVIACEDTGDID-VPCAFNDRKTAYAESFRYLKSRGHENIAF 179

Query: 181 LFSRNNESSATYQSALLAYQEYVQQLSSPYMVGNVHDFNDG-LNLSYQLVKEVSIDGIL 239  
 R + S + AY+ V G+L +M+ G +D NDG L + + I

5 Sbjct: 180 TCVREADRSPSTADKAAAYKAVCGRLDRHMLSG-CNDMNDGELAAEHFYMSGRVPTAIY 238

Query: 240 ATSDEVAAGLIKGYEESRKKCPYIIGQECLLVGQLLKLPTIDHKSYYLGKLAFAKQALAEK 299  
 A SDEVAAG I + + IIG+ + ++L P++D LG AF L ++

10 Sbjct: 239 ANSDEVAAG-IHLFAKNNWVDEIIGEGNTSISRVLGFPSLDLNLLEQLGIAAFSLFLQDE 297

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2171

A DNA sequence (GBSx2288) was identified in *S.agalactiae* <SEQ ID 6707> which encodes the amino acid sequence <SEQ ID 6708>. Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3451(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:AAC21682 GB:U32686 conserved hypothetical protein [Haemophilus influenzae Rd]  
 Identities = 79/264 (29%), Positives = 134/264 (49%), Gaps = 16/264 (6%)

Query: 1 MTIKRIFCDMDGTLNSEQVSKSNATLIREAA---IPVTLVSARAPMEMKDAVDALQLG 57  
 M K +F D +GTLL S+ +S +I+ IP +SAR+P+ + L+

30 Sbjct: 1 MMYKAVFSDFNGLTLLTSQHTISPERTVVVIKRLTANGIPFVPI SARSPGLGILPYWKQLETN 60

Query: 58 GVQVAFNGGLIYRIGDNNQVLPHTQIIEKSTVKQLLRGIRFHFQVSLSYDNLNNWYCD 117  
 V VAF+G LI N + PI++ I+ + ++ + H P + ++YY N+ +

35 Sbjct: 61 NVLVAFGALIL----NQNLPIYSVQIEPKDILEINTVLAEH-PLLGVNYYTNNDCCHAR 115

Query: 118 KID-EGIRYEHSLTQQCPTFIHNEDQFLEGHINTFKIMMITFDEANMLELEKYLSLELP 176  
 ++ + + YE S+T+ IH D+ T + + I + ++E+E L+ + P

Sbjct: 116 DVENKWVIYERSVTK---IEIHPFDEVA---TRSPHKIQIIGEAEEIIEIEVLLKE-KFP 168

40 Query: 177 EITIQRSGKAYLEITHLLAKKSKGIAYILQKEQLAREETAAFGDGHNDLPMLEMVGYPPIV 236  
 ++I RS +LE+ H A K + ++ + E AFGD NDL MLE VG +

Sbjct: 169 HLSICRSHANFLEV MHKSATKGS AVRFL EDYFGVQTNEVIAFGDNFNDLDMLEHVGLGVA 228

Query: 237 MDNAFDDIKAIAYQLTKSNDEDEGV 260  
 M NA ++IK A +T +N+EDG+

45 Sbjct: 229 MGNAPNEIKQAANVVTATNNEDGL 252

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2172

A DNA sequence (GBSx2289) was identified in *S.agalactiae* <SEQ ID 6709> which encodes the amino acid sequence <SEQ ID 6710>. Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2854(Affirmative) < succ>



-2447-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2173

10 A DNA sequence (GBSx2290) was identified in *S.agalactiae* <SEQ ID 6711> which encodes the amino acid sequence <SEQ ID 6712>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have an uncleavable N-term signal seq

|    |          |                     |               |                        |
|----|----------|---------------------|---------------|------------------------|
| 15 | INTEGRAL | Likelihood = -10.51 | Transmembrane | 392 - 408 ( 376 - 417) |
|    | INTEGRAL | Likelihood = -9.92  | Transmembrane | 440 - 456 ( 433 - 461) |
|    | INTEGRAL | Likelihood = -6.42  | Transmembrane | 52 - 68 ( 51 - 70)     |
|    | INTEGRAL | Likelihood = -6.32  | Transmembrane | 29 - 45 ( 9 - 48)      |
|    | INTEGRAL | Likelihood = -6.32  | Transmembrane | 309 - 325 ( 308 - 328) |
|    | INTEGRAL | Likelihood = -4.46  | Transmembrane | 12 - 28 ( 9 - 29)      |
| 20 | INTEGRAL | Likelihood = -3.29  | Transmembrane | 463 - 479 ( 462 - 479) |
|    | INTEGRAL | Likelihood = -2.07  | Transmembrane | 353 - 369 ( 352 - 369) |
|    | INTEGRAL | Likelihood = -1.17  | Transmembrane | 374 - 390 ( 374 - 390) |
|    | INTEGRAL | Likelihood = -0.85  | Transmembrane | 247 - 263 ( 247 - 263) |
|    | INTEGRAL | Likelihood = -0.06  | Transmembrane | 278 - 294 ( 278 - 294) |

25 ----- Final Results -----

bacterial membrane --- Certainty=0.5203 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]  
 Identities = 325/705 (46%), Positives = 478/705 (67%), Gaps = 3/705 (0%)

|    |            |                                                                   |
|----|------------|-------------------------------------------------------------------|
| 35 | Query: 1   | MLQLTKYFPLKPIYLALLVFQIYLLVFSWTMLGCAFLFSFIFLIYQYDRETIFKTIAIV 60    |
|    |            | MLQ K F + IYL+ L+ +Y +FS + L +F + L Q+ ++ K + I                   |
|    | Sbjct: 1   | MLQWIKNFISIPLIYLSFLLWLYYAIFSASYLALLGFVFLVCLFIQFPWKSAGKVLIIIC 60   |
| 40 | Query: 61  | IFFLFYFLWQNHNMNVQYQVRPNHISQIKVRIDTISINGDVLFSQADASGNTYQAFYTLK 120  |
|    |            | F F+F++QN + Q + + + + + + DT+ +NGD LSF+ A G +Q +Y L+              |
|    | Sbjct: 61  | GIFGFWFVFQNWQSQSQSNLADSVRVRILEPDTVKVNGDSLFRGKADGRIFQVYKQLQ 120    |
| 45 | Query: 121 | NKSEKDYFQNLDDNNIMIADIKLEAEERRHFNGFDYRQYLKRHGIYRIAKVTIKIQIRL 180   |
|    |            | ++ EK+ FQ L + I + KL E E +R+F GF+Y+ YLK GIY+ + KI+ ++             |
|    | Sbjct: 121 | SEEEKFAFQALTDLHEIGLEGKLSPEEGQRNFGGFNYQAYLKTQGIYQTLNKKIQSLQK 180   |
| 50 | Query: 181 | FQHRSFALMSKWRSSAIVISQT-FPNPMRHYMSGLLFGYLDKTFDDMSDLYSSLGIIHL 239   |
|    |            | +S RR A+V +T FP+PMR+YM+GLL G+LD F++M++LYSSLGIIHL                  |
|    | Sbjct: 181 | IGSWDIGENLSSLRKAVVWIKTHFPDPMRNYMTGLLLGHLDTFEEMNELYSSLGIIHL 240    |
| 55 | Query: 240 | FALSGMQVGFFLGIYFRYICRLGRDLHVWLLQIPFSLIYAGLTGFSISVVRALIQSLLS 299   |
|    |            | FALSGMQVGFF+ F+ + LR+GL + + L PFSLIYAGLTGFS SV+R+L+Q LL+          |
|    | Sbjct: 241 | FALSGMQVGFFMNGFCKLLRLGLTQEKLKWLTYPFSLIYAGLTGFSASVIRSLQLKLLA 300   |
|    | Query: 300 | HSGVKKDDENFALCLLICISLPHSLLTGGVLSFAYAFILTMSTFDHFSSIKKVAIESLT 359   |
|    |            | GVK +N AL +L+ I +P+ T GGVLS AYAFILTM S + +K VA ESL                |
|    | Sbjct: 301 | QHGVKGLDNCALTIVLVLFIVMPNFFPTAGGVLSCAYAFILTMPSKEG-EGLKAVAGESLV 359 |
|    | Query: 360 | VSVGILPILTYFSGFQFISIIILTALLSFAFDIIFLPLLTIVFVLSPIVKLSCINSLEFI 419  |
|    |            | +S+GILPIL++YF+ FQP SI+LT + SF FD+ FLPLL+++FVLS + + +N +FE         |

-2448-

Sbjct: 360 ISLGILPILSFYFAEFQPSILLTFVFSFLFDLTFLLSILFVLSFLYPVIQLNFIFEW 419

Query: 420 LEVLLKWTGQLFPRPLIFGKPSLFLIVMIIILGLLYDYHSCFRYCSLLIIFTLFFIT 479  
LE +++ Q+ RPL+FG+P+ +LLI+++I L L+YD + L+I LF +T

5 Sbjct: 420 LEGIIRLVSQVTSRPLVFGQPNTWLLILLISLALVYDLRKNIKKLTIVLCLLITGLFLLT 479

Query: 480 KNPITNEVAILDVGQGSILVRDWLGKTLIDTGGRVR-FEQPEEWKQKVNQSNAKRTLII 538  
K+P+ NE+ +LDVGQG+SI +RD GKTLID GG+ +++ ++W++K+ SNA+R+LI

10 Sbjct: 480 KHPLENEITMLDVGQGESIFLRDVTGKTLIDVGGKAESYKKIKKWQEKMTTSNAQRSLI 539

Query: 539 PYLKSRGISKIDDLVITHDTDHMGDMEVISKHFKVARLITSSGSLTNSQYVKHLSKIGV 598  
PYLKSRG++KID L++T+TD +H+GD+ ++K F V ++ S SL ++V L

Sbjct: 540 PYLKSRGVAKIDQLILTNTDKEHVGDLSMTKAFHVGEILVSKDSLKQKEFVAELQATQT 599

15 Query: 599 AVKSIEAGDKLAVMGSYLQVLPWHKGDGKNNDSTVLYGHLLGKGFLLFTGDLEEEGEKQL 658  
V+S+ G+ L + GS L+VL P GDG ++D++VLYG L K FLFTG+LEE+GEK L

Sbjct: 600 KVRSMIVGENLPFGSQLEVLSPRKMGGHDDTLVLYGKFLDKQFLFTGNLEEKGEKDL 659

Query: 659 LEAYPNLSVDILKAGHHGSGKSSSLFLKKLSPSVVLVSAGKNNR 703  
L+ YP+L V++LKA HG+K SSS +FL+KL P + L+S GK+NR

20 Sbjct: 660 LKHYPDLKVNVLKASQHGKSSSPAFLEKLKPELTLISVGKSNR 704

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6713> which encodes the amino acid sequence <SEQ ID 6714>. Analysis of this protein sequence reveals the following:

25 Possible site: 29  
>>> Seems to have an uncleavable N-term signal seq

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.19 | Transmembrane | 394 - 410 ( 380 - 422) |
| INTEGRAL | Likelihood = -8.28  | Transmembrane | 54 - 70 ( 52 - 72)     |
| INTEGRAL | Likelihood = -6.32  | Transmembrane | 356 - 372 ( 355 - 377) |
| INTEGRAL | Likelihood = -4.73  | Transmembrane | 8 - 24 ( 7 - 25)       |
| INTEGRAL | Likelihood = -4.30  | Transmembrane | 30 - 46 ( 29 - 50)     |
| INTEGRAL | Likelihood = -3.88  | Transmembrane | 249 - 265 ( 249 - 267) |
| INTEGRAL | Likelihood = -3.40  | Transmembrane | 467 - 483 ( 465 - 484) |
| INTEGRAL | Likelihood = -2.39  | Transmembrane | 325 - 341 ( 325 - 347) |
| INTEGRAL | Likelihood = -0.43  | Transmembrane | 441 - 457 ( 441 - 458) |

30

35

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5076(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

40

The protein has homology with the following sequences in the databases:

>GP:AAC23742 GB:AF052208 competence protein [Streptococcus pneumoniae]  
Identities = 311/706 (44%), Positives = 458/706 (64%), Gaps = 10/706 (1%)

45 Query: 5 WTKLVPLSKIQFAFLILVFFYQIHSPSWLTFL-LSLSLICLLVKRLSKK--EFLGVFAIL 61  
W K + I +FL+L +Y I S S+L L L+CL ++ K + L + I

Sbjct: 4 WIKNFSIPLIYLSFLLWLWYIAIFSASYLALLGFFVLLVCLFIQFPWKSAGKVLIICGIF 63

50 Query: 62 SFCALFLLYQKQQLVQKLEIQPVQITSVALVPDSIRINGDQLAVLGRHGKHSYQLFYRLK 121  
F +F +Q+ Q Q L + V ++PD++++NGD L+ G+ +Q++Y+L+

Sbjct: 64 GFWFVFQNWQQSQASQNLADS---VERVRILPDTVKVNGDSL SFRGKADGRIFQVYKQL 120

55 Query: 122 SQAEALFKKEHRWLVHAKVITLKAEEVRNFKGFNYQTFLTYQGIYRIGKVEQIEQLEV 181  
S+ E + F+ + + L + E RNF GFNYQ +L QGIY+ +++I+ L+

Sbjct: 121 SEEEKAEAFQALTDLHEIGLEGLKSEPEGQRNFGGFNYQAYLKTQGIYQTLNKKIQSLQK 180

Query: 182 ISPESICDYLLSLRRRAIVHCQQHFPMPMSHYLTGLLFGYLDKSFGEIMTDYYSQLGIIHL 241  
I I + LSSLRR+A+V + HFP PM +Y+TGLL G+LD F EM + YS LGIIHL

60 Sbjct: 181 IGSDWIDIGENLSSLRRKAVVWIKTHFPDPMRNYMTGLLLGHLDTFEEMNELYSSLGIIHL 240

Query: 242 FALSGMQVGFFLTFCFRVLLLLAVPLEWIKWIELPFACFYAALTGYSISVIRSLVQSOLR 301  
FALSGMQVGFF+ F+++LL L + E +KW+ PF+ YA LTG+S SVIRSL+Q L

65 Sbjct: 241 FALSGMQVGFFMNGFKKLLLRGLTQEKLKWLTYPFSLIYAGLTGFSASVIRSLQLKLLA 300

-2449-

Query: 302 HLGKGLDNLACTFLLVFLWDAHFLMTVGGVLTFSYAFLLTVVTVEELSGAKRQLVQVLT 361  
 G+KGLDN A T L++F+ +F T GGVL+ +YAF+LT+ + +E G K + L

Sbjct: 301 QHGVKGLDNCALTVLVLFIVMPNFFFTAGGVLSCAYAFILTMPS-KEGEGLKAVASESLV 359

5 Query: 362 ISLGILPFLLFYFSSFNPMMSVLTGLLSYLFDLFILPLLCLVFCLSPLVTVSICNHLFIL 421  
 ISLGILP L FYF+ F P S++LT + S+LFDL LPLL ++F LS L V N +F

Sbjct: 360 ISLGILPILSFYFAEFQPSIILLTFVFSFLFDLTFLLSILFVLSFLYPVIQLNFIFEW 419

10 Query: 422 LEKVIQFLGNTFNSSLVFGSPTSWHLLILVISFAIFYDYRQ-VRQRVITCGLVIALTLLS 480  
 LE +I+ + + LVFG P +W L++L+IS A+ YD R+ +++ + C L+ L LL+

Sbjct: 420 LEGIIRLVSVQTSRPLVFGQPNTWLLILLISLALVYDLRKNIKLTVLCLLITGLFLLT 479

15 Query: 481 VKYPLTNEVTFIDIGQDSILVREWTGKNLLIDVGGR-PFFSSKEHWRRGHVANAQKTL 539  
 K+PL NE+T +D+GQG+SI +R+ TGK +LIDVGG+ + + W+ +NAQ++L

Sbjct: 480 -KHPLENEITMLDVGQGESIFLRDVTGKTILIDVGGKAESYKKIKKQEKMTTSNAQRSL 538

20 Query: 540 IPYLKSRGIHTIDQLLVTHADTDHMGDIEVVAKAIRIKEILTSQGSLSHPSFVRRLRLK 599  
 IPYLKSRG+ IDQL++T+ D +H+GD+ + KA + EIL S+ SL FV L+ +

Sbjct: 539 IPYLKSRGVAKIDQLILNTDKEHVGLDSEMTKAFHVGEILVSKDSLKQKEFVAELQATQ 598

25 Query: 600 CHVRVLAAGDQLPIMGSVLQVLYPWQLGDKNNDSLVLYGRLLNRTFLFTGDLEKEGENE 659  
 VR + G+ LPI GS L+VL P ++GDG ++D+LVLYG+ L++ FLFTG+LE++GE +

Sbjct: 599 TKVRSMIVGENLPFIGSQLEVLSPRKMGGHDDTLVLYGKFLDKQFLFTGNLEEKGEKD 658

Query: 660 IIKRYPQLRVDYLKAGHHSNTSSSAFLDHIQPKVAFISAGKNNR 705  
 ++K YP L+V+ LKA HG+ SSS AFL+ ++P++ IS GK+NR

Sbjct: 659 LLKHYPDLKVNVLKASQHGKSSSPAFLEKLKPELTLISVGKSNR 704

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 346/743 (46%), Positives = 491/743 (65%), Gaps = 3/743 (0%)

Query: 5 TKYFPLKPIYLALLVFQIYLLVFSWTLGCAFLLESFIFLIYQYDRETIFKTI AIVIFFL 64  
 TK PL I A L+ + + S + L L L+ + ++ AI+ F

35 Sbjct: 6 TKLVPLSKIQFAFLILVFFYQIHSPSWLTFLLSLSLICLLVKRLSKKEFLGVFAILSFC 65

Query: 65 FYFLWQNHNMNVQYQRPVNHISQIKVRIDTISINGDVL SFQADASGNTYQAFYTLKNKSE 124  
 + L+Q + + + P I+ + + D+I INGD L+ ++YQ FY LK+++E

Sbjct: 66 LFLLYQKQQLVQKLEIQPVQITSVALVPDSIRINGDQLAVLGRHGKHSYQLFYRLKSQAE 125

40 Query: 125 KDYFQNLNMMIADIADIKLEEAERRHFNDFYRQYLRHGIYRIAKVTIKIQIRLFQHR 184  
 F+ +++ A + LE+AEE R+F GF+Y+ +L GIYRI KV +I+Q+ +

Sbjct: 126 AQLFKKEHRWLVMAKVTLEKABEVRNFKGFNYQTFITYQGIYRIGKVEQIEQLEVISPE 185

45 Query: 185 SFFALMSKWRRSAIV-ISQTFPNPMRHYSGLLFGYLDKTFDDMSDLYSSLGIIHLFALS 243  
 S +S RR AIV Q FP PM HY++GLLFGYLDK+F +M+D YS LGIIHLFALS

Sbjct: 186 SICDYLSLRRRAIVHCQQHFPRPMSHYLTGLLFGYLDKSFGEEMTDYYSQLGIIHLFALS 245

50 Query: 244 GMQVGFFLGIFRYICRIGRLDHVWLLQIPFSLIYAGLTGFSISVVRALIQSLLSHSGV 303  
 GMQVGFFL FR + L + + L+ + +++PF+ YA LTG+SISV+R+L+QS L H G+

Sbjct: 246 GMQVGFFLTCFRRVLLLLLAVPLEWIKWIELPFACFYAALTGYSISVIRSLVQSQLRHLGI 305

55 Query: 304 KKDENFALCLLILCLISLPHSLTTGGVLSFAYAFILTMSTFDHFSSIKKVAIESLTIVSVG 363  
 K +N A L+ + H L+T GGVL+F+YAF+LT+ + + S K+ ++ LT+S+G

Sbjct: 306 KGLDNLACTFLLVFLWDAHFLMTVGGVLTFSYAFLLTVVTVEELSGAKRQLVQVLTISLG 365

60 Query: 364 ILPILTYFSGFQIPISIIITALLSFAFDIIFLPLLTVIFVLSPIVKLSCINSLEIFEILEVL 423  
 ILP L +YFS F P+S++LT LLS+ FD+ LPLL ++F LSP+V +S N LF +LE +

Sbjct: 366 ILPFLLFYFSSFNPMMSVLTGLLSYLFDLFILPLLCLVFCLSPLVTVSICNHLFILLEKV 425

65 Query: 424 LKWTGQLFPRPLIFGKPSLFLIVMIIILGLLYDYHYSKC-FRYCSLLIIFTLFFITKNP 482  
 +++ G F L+FG P+ + L++++I + YDY + C L+I TL + K P

Sbjct: 426 IQFLGNTFNSSLVFGSPTSWHLLILVISFAIFYDYRQVRQRVITCGLVIALTLLSV-KYP 484

Query: 483 ITNEVAILDVGQDSILVRDWLGKTLIDTGGRRVFEQPEEWKQKVNQSNKRTLIPYLK 542  
 +TNEV +D+GQDSILVR+W GK +LID GGR F E W++ + +NA++TLIPYLK

Sbjct: 485 LTNEVTFIDIGQDSILVREWTGKNLLIDVGGRPFFSSKEHWRRGHVANAQKTLIPYLK 544

-2450-

Query: 543 SRGISKIDDLVITHDTDTHMGDMVISKHFVARLITSSGSLTNSQYVKHLSKIGVAVKS 602  
 SRGI ID L++TH DTDHMGD+EV++K ++ ++TS GSL++ +V+ L ++ V+  
 Sbjct: 545 SRGIHTIDQLLVTHADTDHMGDIEVAKAIRIKEILTSQGSLSHPSFVRLRLRLKCHVRV 604

5 Query: 603 IEAGDKLAVMGSYLQVLYPWHKGDGKNDSIVLYGHLLGKGLFTGDLEEEGEKQLLEAY 662  
 + AGD+L +MGS LQVLYPW GDGKNDS+VLYG LL + FLFTGDLE+EGE ++++ Y  
 Sbjct: 605 LAAGDQLPIMGSVLQVLYPWQLGDGKNDSLVLYGRLLNRTFLFTGDLEKEGENEIIKRY 664

10 Query: 663 PNLSDILKAGHHGSKGSSSLFLLKLSPSVVLVSAGKNNRYQHPHQETLQRFQIKISKI 722  
 P L VD LKAGHHGS SSS +FL + P V +SAGKNNRYQHPH+ETL R + +  
 Sbjct: 665 PQLRVDYLKAGHHGSNTSSSAFLDHIQPKVAFISAGKNNRYQHPHRETLLARLEDRQITY 724

Query: 723 FRTDQSGTIRLTGWKWHIQTVR 745  
 +RTD G IRLTG WH++TVR  
 15 Sbjct: 725 YRTDTQGAIRLTGRTSWHLETVR 747

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2174

20 A DNA sequence (GBSx2291) was identified in *S.agalactiae* <SEQ ID 6715> which encodes the amino acid sequence <SEQ ID 6716>. This protein is predicted to be competence protein (comEA). Analysis of this protein sequence reveals the following:

Possible site: 38  
 >>> Seems to have an uncleavable N-term signal seq  
 25 INTEGRAL Likelihood = -3.77 Transmembrane 18 - 34 ( 14 - 36)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2508(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]  
 Identities = 96/217 (44%), Positives = 138/217 (63%), Gaps = 4/217 (1%)

35 Query: 3 EIVLEKIKSHKWETTGIIIVGLLLFGILGLNHFH-TTHKEDNLNINLEK-KVSTITEKKVP 60  
 E ++EKIK +K +GLL+ G L T KE NL + ++EK+V  
 Sbjct: 2 EAIIKEIKKEYKIIIVICTGLGLLVGGFLLKPAQTPVKETNLQAEVAAVSKDLVSEKEVN 61

40 Query: 61 MISHVKDKVSNQVTVDVKGAVNHPGVYSLPSQSRVTDKIKRAGGLSNLADSKSVNLAQKL 120  
 + + +TVDVKGAV PG+Y LP SR+ DA+++AGGL+ ADSKS+NLAQK+  
 Sbjct: 62 KEEKEEPLEQDLITVDVKGAVKSPGIYDLPVGSRIINDAVQKAGGLTEQADSKSLNLAQKV 121

45 Query: 121 QDETVIYVAQKGEKITVVEEEKANNIATQGNSKGKINLNKADLSSLQTISGVGAKRAQDI 180  
 DE ++YV KGE+ V ++ A+ + + K+NLNKA L L+ + G+G KRAQDI  
 Sbjct: 122 SDEALVYVPTKGEE--AVSQQTGLGTASSISKEKKVNLNKASLEELKQVKGLGGKRAQDI 179

Query: 181 LDYRDSQGGFKTIDDLKNVSGIGEKTEKLRQDVTID 217  
 +D+R++ G FK++D+LK VSGIG KT+EKL+ VT+D  
 50 Sbjct: 180 IDHREANGKFKSVDELKKVSGIGGKTIEKLDYVTVD 216

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6717> which encodes the amino acid sequence <SEQ ID 6718>. Analysis of this protein sequence reveals the following:

Possible site: 36  
 55 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -9.61 Transmembrane 22 - 38 ( 16 - 42)

----- Final Results -----

-2451-

bacterial membrane --- Certainty=0.4843(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:AAC23741 GB:AF052208 competence protein [Streptococcus pneumoniae]  
 Identities = 82/179 (45%), Positives = 124/179 (68%), Gaps = 4/179 (2%)

10 Query: 42 NRQSKAAVPALREISPVKQVSEKKKEIQEDSSILVDLKGAVQKEGVYKLTASSRVRDVI 101  
 N Q++ A + +++ K+ EEK+E E I VD+KGAV+ G+Y L SR+ D +  
 Sbjct: 42 NLQAEVAAVS-KDLVSEKEVNKEEKEEPLQDLITVDVKGAVKSPGIYDLFVGSRRINDAV 100

Query: 102 ELAGGLTSEADKHAINFAEKLTDQVVVYPKQGEESVLPRSLVSGKKETASKDQSKVHI 161  
 + AGGLT +AD ++N A+K++DE +VYVP +GEE + + G + SK++ KV++  
 15 Sbjct: 101 QKAGGLTEQADSKSLNLAQKVSDEALVYVPTKGEE--AVSQQTGLGTASSISKEK-KVNL 157

Query: 162 NKASLEELQHPIPGIGAKRAQDIIDMRDKLGGFKALEDLRQVSGIGEKLTLEKLKDDIFLD 220  
 NKASLEEL+ + G+G KRAQDIID R+ G FK+++L++VSGIG KT+EKLKD + +D  
 20 Sbjct: 158 NKASLEELKQVKGLGGKRAQDIIDHREANGKFKSVDELKKVSGIGGKTIEKLKDYVTVD 216

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 81/166 (48%), Positives = 111/166 (66%), Gaps = 10/166 (6%)

25 Query: 62 ISHVKDKVSNQ-----VTVDVKGAVNHPGVYSLPSQSRVTDAIKRAGGLSNLADSK 112  
 IS VK +VS + + VD+KGAV GY L + SRV D I+ AGGL++ AD  
 Sbjct: 55 ISPVKQVSEKKKEIQEDSSILVDLKGAVQKEGVYKLTASSRVRDVIELAGGLTSEADKH 114

Query: 113 SVNLAQKLQDETVIYVAQKGEKITVVEEEKANNIA-TQNSKGKINLNKADLSSLQTISG 171  
 ++N A+KL DE V+YV ++GE+I+V+ + T + K+++NKA L LQ I G  
 30 Sbjct: 115 AINFAEKLTDQVVVYPKQGEESVLPRSLVSGKKETASKDQSKVHINKASLEELQHPIG 174

Query: 172 VGAKRAQDILDYRDSQGGFKTIDDLKNVSGIGEKLTLEKLQDVTID 217  
 +GAKRAQDI+D RD GGFK ++DL+ VSGIGEKLTLEKL+ D+ +D  
 35 Sbjct: 175 IGAKRAQDIIDMRDKLGGFKALEDLRQVSGIGEKLTLEKLKDDIFLD 220

A related GBS gene <SEQ ID 8989> and protein <SEQ ID 8990> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 9  
 McG: Discrim Score: 5.70  
 GvH: Signal Score (-7.5): -2.58  
 Possible site: 38  
 >>> Seems to have an uncleavable N-term signal seq  
 ALOM program count: 1 value: -3.77 threshold: 0.0  
 INTEGRAL Likelihood = -3.77 Transmembrane 18 - 34 ( 14 - 36)  
 45 PERIPHERAL Likelihood = 10.40 73  
 modified ALOM score: 1.25

\*\*\* Reasoning Step: 3

50 ----- Final Results -----

bacterial membrane --- Certainty=0.2508(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

44.3/64.1% over 215aa

Streptococcus

pneumoniae

GP|3211753| competence protein Insert characterized

60 ORF01930(304 - 951 of 1014)

GP|3211753|gb|AAC23741.1|AF052208(1 - 216 of 216) competence protein {Streptococcus pneumoniae}

-2452-

```
%Match = 25.0
%Identity = 44.2 %Similarity = 64.1
Matches = 96 Mismatches = 75 Conservative Sub.s = 43
```

```

5      90          120          150          180          210          240          270          300
DDGKKLNPLTYIYRLPLAIIAIVLLVLTILFSYLASFVWDPQKHLK*GLHGNLYLLFSK*FFWFLIGKSL*LRISKWRNIF

330          360          390          417          447          474          504          534
MFEIVLEKIKSHKWETTGIIVGLLLFGILGLNHFG-THHKEDNLNINLEK-KVSTITEKKVPMISHVKDKVSNQVTVVVK
10    |::||| |:| :|||: |:| ||||: :|||:| : :|||
MEAIIEKIKEYKIIIVICTGLGLLVGGFFLLKPAPQTPVKE'TNLQAEVAAVSKDLVSEKEVNKEEKEEPLEQDLITVDVK
                10          20          30          40          50          60          70

564          594          624          654          684          714          744          774
GAVNHPGVYSLPSQSRSVTDAIKRAGGLSNLADSKSVNLAQKLQDETVIYVAQKGKITVVXEEKANNIATQGNSKSGKINL
15    ||| ||:| ||| :||::|||: ||||:||||: ||::| |||: |:: |: : :|:|
GAVKSPGIYDLFVGSRINDAVQKAGGLEQADSKSLNAQKVSDREALVYPVP'KGE--AVSQQTGLGTASSISKEKKVNL
                90          100          110          120          130          140          150

804          834          864          894          924          954          984          1014
NKADLSSLQTIISGVGAkraQDILDYRDSQGGFKTIDDLKNVSGIGEKTLEKLRQDVTTID*VFSSKTYLFSIVGLPNLLTS
20    ||| | |: : |:| |||||:|: ::| ||:::| |||| | |:| |:|
NKASLEELKQVKGGLGGKRAQDIIDHREANGKFksvDELKKVSGIGGKTIEKLDKYVTVD
                170          180          190          200          210

```

25 SEQ ID 8990 (GBS129) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 4; MW 43.8kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2175

30 A DNA sequence (GBSx2292) was identified in *S.agalactiae* <SEQ ID 6719> which encodes the amino acid sequence <SEQ ID 6720>. Analysis of this protein sequence reveals the following:

Possible site: 54  
>>> Seems to have a cleavable N-term signal seq.  
INTEGRAL Likelihood =-14.01 Transmembrane 215 - 231 ( 208 - 240)

```

35      ----- Final Results -----
          bacterial membrane --- Certainty=0.6604 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12793 GB:Z99109 similar to 1-acylglycerol-3-phosphate
O-acyltransferase [Bacillus subtilis]
Identities = 66/200 (33%), Positives = 111/200 (55%), Gaps = 10/200 (5%)
```

45

Query: 3 YTYLRTLVMFLIWVANGNAHYHNEDKMLKDDENIYLVAHRTFWDPVYMAFAARPKQFIF 62  
Y + + ++ G Y+ E+ L D +++ H + D + + P Q +  
Sbjct: 2 YKFCANALKVILSLRGGVKVYNKEN--LPADSGFVIACHTSGWVDVITLGVGILPYQIH 59

50 Query: 63 MAKKELFTNRLFGWWMKCGAFPIDREKPGQDAIRYPVKMLKNSNRLVMFPGSGSRHSD 122  
MAKKELF N+ G ++K AFGDRE PG +I+ P+K+LK + +FPGS+R S+D  
Sbjct: 60 MAKKELFQNKWIGSFLKKIHAFVDRENPGPSSIKTPIKLLK-EGEIVGIFPGS+RTS+D 118

55 Query: 123 V--KGGVAVIAKMAKVRIMPAAYRGPVFKNLLKGHRVDMNFGNPIDVSDIKRMDA-EGI 179  
V K G IA+M K ++PAAY+GP K L K ++ + G P+ +D + + E +  
Sbjct: 119 VPLKRGAVTTIAQMKGAPLVPAAAYQGPSSGKELFKKGKMKLIIGEPLHQADFALHPSKERL 178

Query: 180 A---EVSRRIQEEFDRLDR 195  
A +++RI+E ++LD+  
60 Sbjct: 179 AAMTEALNQRIKELENKLDQ 198

:2453-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6721> which encodes the amino acid sequence <SEQ ID 6722>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.83    Transmembrane    241 - 257 ( 234 - 266)
    INTEGRAL    Likelihood = -4.41    Transmembrane    27 - 43 ( 26 - 44)

----- Final Results -----
10      bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

15  >GP:CAB12793 GB:Z99109 similar to 1-acylglycerol-3-phosphate
      O-acyltransferase [Bacillus subtilis]
      Identities = 59/198 (29%), Positives = 104/198 (51%), Gaps = 6/198 (3%)

20  Query: 29 YAYLRLGLVVFLLWVNGNAHYHHEEKMLDASENYILVAPHRTFWDPVYMAFAARPKQFIF 88
      Y + + +L+ G Y+ E L A +++ H + D + + P Q +
      Sbjct: 2 YKFCANALKVILSLRGGVKVYNKEN--LPADSGFVIACHTSGWVDVITLGVGILPYQIH 59

      Query: 89 MAKKELFANRLFPAWWIKMCGAFFIDRDKPSPDAIRYPVNMLKKSNNRSLMFPSGSRHSQE 148
      MAKKELF N+ ++K AFP+DR+ P P +I+ P+ +LK+ + +FPSG+R S++
25  Sbjct: 60 MAKKELFQNKWIGSFLKKIHAFPPVDRENPGPSSIKTPIKLLKE-GEIVGIFPSGTRTSED 118

      Query: 149 V--KGGVAVIAKLAKVKIMPAAYQGPM SVKGLLAGERVDMTFCGNPIDVSDIKRM-NDEGI 205
      V K G IA++ K ++PAAYQGP S K L ++ + G P+ +D + + E +
30  Sbjct: 119 VPLKRGAVTIAQMKGAPLVPAAYQGPPSSGKELFKKGKMKLIIGEPLHQADFAHLPSKERL 178

      Query: 206 AEVANRIQAEFDRIDDEL 223
      A + + ++++L
      Sbjct: 179 AAMTEALNQRIKELNKL 196

```

35 An alignment of the GAS and GBS proteins is shown below.

```

      Identities = 186/244 (76%), Positives = 212/244 (86%)

40  Query: 1 MFYTYLRLTLVLMFLIWVANGNAHYHNEEDKMLKDDENYILVAPHRTFWDPVYMAFAARPKQF 60
      +FY YLR LV+FL+WV NGNAHYH+E+KML ENYILVAPHRTFWDPVYMAFAARPKQF
      Sbjct: 27 VFYAYLRLGLVVFLLWVNGNAHYHHEEKMLDASENYILVAPHRTFWDPVYMAFAARPKQF 86

      Query: 61 IFMAKKELFTNRLFGWWIKMCGAFFIDREKPGQDAIRYPVKMLKNSNRLVMPFSGSRHS 120
      IFMAKKELF NRLF WWIKMCGAFFIDR+KP DAIRYPV MLK SNRSL+MPFSGSRHS
45  Sbjct: 87 IFMAKKELFANRLFPAWWIKMCGAFFIDRDKPSPDAIRYPVNMLKKSNNRSLMFPSGSRHS 146

      Query: 121 KDVKGGVAVIAKMAKVRIMPAAYRGPMVFKNLLKGHRVDMNFGNPIDVSDIKRMDAEGIA 180
      ++VKGGVAVIAK+AKV+IMPAAY+GPM K LL G RVDM FGNPIDVSDIKRM+ EGIA
      Sbjct: 147 QEVKGGVAVIAKLAKVKIMPAAYQGPM SVKGLLAGERVDMTFCGNPIDVSDIKRMNDEGIA 206

50  Query: 181 EVSRRIQEEDRLDRENETYDDGKKLNPLTYIYRLPLAIIAIVLLVLTILIFSYLASFVWD 240
      EV+ RIQ EFDR+D E + GK NPLTY+YRLPL ++ +V+L+LT++FSY+ASFVW+
      Sbjct: 207 EVANRIQAEFDRIDDELAPFQPGKARNPLTYLYRLPLGLVLVVVLLLTMLFSYIASFVWN 266

55  Query: 241 PQKH 244
      P KH
      Sbjct: 267 PDKH 270

```

SEQ ID 6720 (GBS171) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 36 (lane 2; MW 25kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 3; MW 49.8kDa).

-2454-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2176**

A DNA sequence (GBSx2293) was identified in *S.agalactiae* <SEQ ID 6723> which encodes the amino acid sequence <SEQ ID 6724>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.3268(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15   >GP:CAB11810 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]  
      Identities = 113/244 (46%), Positives = 173/244 (70%), Gaps = 2/244 (0%)

Query: 6   LKENERIDQLFSTDVKIIQNKEVFSYSIDSVLLSRFPKLP-SRGLIVDLCGNGAVGLFA 64  
          L ++ER+D L + D+KIIQ+ VF++S+D+VLLS+F +P +G IVDLC+GNG V L  
20   Sbjct: 4   LHDDERLDYLLAEDMKIIQSPTVFAFSLDAVLLSKFAYVPIQKGKIVDLC+GNGIVPLLL 63

Query: 65   STKTNATIIIEIQLQESLADMAKRSIKLNKLEKQVTMINDDLKNLLDHVQRSNVDLMLCNP 124  
          ST++ A I+ +E+QE L DMA RS++ NKL+ Q+ +I+DDLKN+ + + + D++ CNP  
25   Sbjct: 64   STRSKADILGVEIQERLHDMAVRSVEYNKLDDQIQIHHDDLKNMPEKLGHNRYDVVTCNP 123

Query: 125   PYFKASETSKKNLSPHYLLARHEITTNLREICQIAQHALKTKGRIAMVHRPDRFLEIIDT 184  
          PYFK + +++N++ H +ARHEI L ++ ++ LK G+ A+VHRP R LEI +  
30   Sbjct: 124   PYFKTPKQTEQNMNEHLRIARHEIHCTLEDVISVSSKLLKQGGAALVHRPGRLLLEIFEL 183

Query: 185   MRQFNLAPKRIQFVYPKLGKDANMLLIEAIKDGSTEGMKILPPLVHVQDNGDYTETIFDI 244  
          M+ + + PKR+QFVYPK GK+AN +L+E IK G + +KILPPL V+ + +YT+ I I  
35   Sbjct: 184   MKAYQIEPKRVQFVYPKQGKEANTILVEGIKGGRPD-LKILPPLFVYDEQNEYTKERTI 242

Query: 245   YFGE 248  
          +G+  
35   Sbjct: 243   LYGD 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6725> which encodes the amino acid sequence <SEQ ID 6726>. Analysis of this protein sequence reveals the following:

40   Possible site: 48

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45           bacterial cytoplasm --- Certainty=0.2183(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50   Identities = 200/257 (77%), Positives = 228/257 (87%), Gaps = 3/257 (1%)

Query: 1   MIDTILKENERIDQLFSTDVKIIQNKEVFSYSIDSVLLSRFPKLP+SRGLIVDLCGNGAV 60  
          MI ILKE ERIDQLFS+DV IIQNK+VFSYSIDSVLLSRFPK+PS+GLIVDLCGNGAV  
55   Sbjct: 1   MIKAILKEGERIDQLFSSDVGIQNKDVFSYSIDSVLLSRFPKMPKGLIVDLCGNGAV 60

Query: 61   GLFASTKTNATIIIEIQLQESLADMAKRSIKLNKLEKQVTMINDDLKNLLDHVQRSNVDLML 120  
          GLFAST+T A I+E+ELQE LADM +RSI+LN+LE QVTMI DDLKNLL+HV RS VDLML  
61   Sbjct: 61   GLFASTRTKAAIVEVELQERLADM+QORSIQLNQLEDQVTIMICDDLKNLLNHVPRSGVDLM 120

Query: 121   LCNPPYFKASETSKKNLSPHYLLARHEITTNLREICQIAQHALKTKGRIAMVHRPDRFLE 180



-2455-

LCNPPYPFK+ E+SKKN+S HYLLARHE+TTNL EICQ+A+HALK+ GR+AMVHRPDRFLE  
 Sbjct: 121 LCNPPYPKSHSSSKKNVSEHYLLARHEVTTNLEEICQVARHALKSNGLAMVHRPDRFLE 180

Query: 181 IIDTMRQFNLAPKRIQFVYPKLGKDANMLLIEAIKDGSTEGMKILPPLVVHQDNGDYTET 240  
 IID++R LAPKR+QFVYPKLGK ANMLLIEAIKDGSG EGM ILPPLVVH++NG+YT+  
 Sbjct: 181 IIDSLRANGLAPKRVQFVYPKLGKSANMLLIEAIKDGSIEGMTILPPLVVHKENGEYTDH 240

Query: 241 IFDIYFGENGK---SHD 254  
 IF+IYFG K +HD  
 Sbjct: 241 IFEIYFGAASKGKPNHD 257

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2177

- 15 A DNA sequence (GBSx2294) was identified in *S.galactiae* <SEQ ID 6727> which encodes the amino acid sequence <SEQ ID 6728>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1512(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11811 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 40/82 (48%), Positives = 63/82 (76%)

Query: 7 YMYVLECSGDTLYTGYYTDDVKRRNLNTHNTGKGAKYTRARLPVKLLYSEAFNSKQEQEAMRAE 66  
 + YV++C D + Y GYT D+ +R+ THN GKGAKYT+ R PV+L+++E+F++K+EAM+AE  
 Sbjct: 7 FFYVVKCKDNSWYAGYTNDLHKRVKTHNDGKGAKYTKVRRPVELIFAESFSTKREAMQAE 66

Query: 67 ALFKQKTRQAKLTYIKQHKEQ 88  
 FK+ TR+ K YI++ +N +  
 Sbjct: 67 YFVKLTKKKELYIEEKRNSK 88

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6729> which encodes the amino acid sequence <SEQ ID 6730>. Analysis of this protein sequence reveals the following:

40 Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1838(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 60/84 (71%), Positives = 67/84 (79%), Gaps = 1/84 (1%)

50 Query: 6 AYMYVLECSGDTLYTGYYTDDVKRRNLNTHNTGKGAKYTRARLPVKLLYSEAFNSKQEQEAMRA 65  
 AYMYVLEC D TLYTGYYTDD+K+RL THN GKGAKYTR RLPV LLY E F+SK+ AM A  
 Sbjct: 6 AYMYVLECDKTLTYTGYYTDDLKRLATHNAGKGAKYTRYRLPVSLLYEYVFDSKEAAMSA 65

Query: 66 EALF-KQKTRQAKLTYIKQHKEQ 88  
 EALF K+KTR KL YI H+ E+  
 Sbjct: 66 EALFKKRKTRSQKLAYIATHQKEK 89

-2456-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2178

A DNA sequence (GBSx2295) was identified in *S. agalactiae* <SEQ ID 6731> which encodes the amino acid sequence <SEQ ID 6732>. This protein is predicted to be autoaggregation-mediating protein (deaD). Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2287(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD20136 GB:AF091502 autoaggregation-mediating protein

[Lactobacillus reuteri]

Identities = 289/504 (57%), Positives = 366/504 (72%), Gaps = 18/504 (3%)

Query: 1 MKFTELNLSDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAGFLPTL 60

MKF+EL LS +L A++++G+ E +PIQE TIP+ LEGKDVIGQAQTGTGKTAAGFLP +

Sbjct: 1 MKFSELGLSDSLKAIKRSGYEATPIQEQTIPMVLEGKDVIGQAQTGTGKTAAGFLPII 60

Query: 61 NKIHTEDNTIQALIIAPTRELAVQSQEELFRFGRDKGVKVRVYGGSSIEKQIKALRSGA 120

+ TE+ IQA+II+PTRELA+Q+QEEL+R G+DK V+V+ VYGG+ I +QIK+L+

Sbjct: 61 ENVDTENPNIQAIISPTRELAQTQEELYRLGKDKHVRVQVYGGADIRRQIKSLKQHP 120

Query: 121 HVVVGTPGRLLDLIKRLKALNHIETLILDEADEMLNMGFLEDIEAIIISRPETRQTLLF 180

++VVGTPGRL D I R +KL+HI+TL+LDEADEMLNMGFLEDIE+II P+ RQTLLF

Sbjct: 121 QILVGTPGRLRDHINRHTVKLDHIKTLVLDEADEMLNMGFLEDIESIIKETPDDRQTLLF 180

Query: 181 SATMPDPIKRIGVKFMKDEPHVKIKATELTNVNVDQYYVRVKENEFDTMTRLMVDVQPE 240

SATMP IKRIGV+FM DPE V+IKA ELT VDQYYVR ++ EKFD MTRL+DV P+

Sbjct: 181 SATMPPEIKRIGVQFMSDPETVRIKAKELTTDLVDQYYVRARDYEKFDIMTRLIDVQDPD 240

Query: 241 LSIVFGRTKRRVDELTRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDHIDILVATDVAA 300

L+IVFGRTKRRVDEL++GL RG+ A GIHGD L Q+KR +++ FKN+ +DILVATDVAA

Sbjct: 241 LTIIVFGRTKRRVDELSKGLIARGYNAAGIHGDLTQDKRSKIMWKFKNNELDILVATDVAA 300

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGQSITFVSPNEMGYLTIENLTCK 360

RGLDISGVTHVYNYDIP DP+SYVHRIGRTGRAG G S+TFV+PNEM YL IE LT+

Sbjct: 301 RGLDISGVTHVYNYDIPSDPDSYVHRIGRTGRAGHHGVSLTFVTPNEMDYLHEIEKLTRV 360

Query: 361 RMTGMKPATASEAFQAKKKVALKRIARDFED-QELVSK--FDKFKADALELATQYTPEEL 417

RM +KP TA EAF+ ++A F D EL+++ D+++ A +L + +L

Sbjct: 361 RMLPLKPPTAEAEAFKG-----QVSAFNDIDELIAQDSTDYEEAAEKLETHNATDL 413

Query: 418 ALYVLSLTVQDPESLPEVEITREKPLPFKPSGGGFKGKGGRGNRGGD--RRRNDRGDRR 475

+L+ ++ S V+IT E+PLP + G R N GG+ RR+N R +

Sbjct: 414 VAALLNNMTKEAASEVPVKITPERPLPRRNKRN--NRNGNRNNSHGGNHYYRRKNFRRHQH 471

Query: 476 GNRDRDDRG----SRCDFKRRDDK 495

G+ D+ G SR F R K

Sbjct: 472 GSHRNDNHGKSHSRHSFNIRHRK 495

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 6733> which encodes the amino acid sequence <SEQ ID 6734>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

-2457-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1108(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 430/545 (78%), Positives = 463/545 (84%), Gaps = 24/545 (4%)

Query: 1 MKFTELNLSQDILSAVEKAGFVEPSPIQEMTIPLALEGKDVIGQAQTGTGKTAAGFLPTL 60  
 +KFTE NLSQDI SAV AGF + SPIQEMTIPLALEGKDVIGQAQTGTGKTAAGFLPTL  
 Sbjct: 1 LKFTEFNLSQDIQSAVVTAGFEKASPIQEMTIPLALEGKDVIGQAQTGTGKTAAGFLPTL 60

Query: 61 NKIHTEDNTIQALIIAPTRELAVQSQEELFRFGRDKGVKVRSVYGGSSIEKQIKALRSGA 120  
 NKI T +N IQAL+IAPTRELAVQSQEELFRFGR+KGVKVRSVYGGSSIEKQIKAL+SGA  
 Sbjct: 61 NKIRTNENIIQALVIAPTRELAVQSQEELFRFGREKGVKVRSVYGGSSIEKQIKALKSGA 120

Query: 121 HVVVGTPGRLLDLIKRKALKLNHIETLILDEADEMLNMGFLEDIEAIIISRPETRTLLF 180  
 H+VVGTPGRLLDLIKRKAL L+H+ETLILDEADEMLNMGFLEDIEAIIISRP RTLLF  
 Sbjct: 121 HIVVGTPGRLLDLIKRKALILDHVETLILDEADEMLNMGFLEDIEAIIISRPADRTLLF 180

Query: 181 SATMPDPIKIGVKFMKDPEHVKIKATELTNNVNDQYYVRVKEKFDTMTRLMVDVQPE 240  
 SATMP PIK+IGVKFMKDPEHV+IK ELTNNVNDQYYVRVKE EKFDTMTRLMDV+QPE  
 Sbjct: 181 SATMPAPIKQIGVKFMKDPEHVQIKNKELTNNVNDQYYVRVKEQKFDTMTRLMVDVNQPE 240

Query: 241 LSIVFGRTKRRVDELTRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDHIDILVATDVAA 300  
 LSIVFGRTKRRVDE+TRGLKLRGFRAEGIHGDLQNKRLRVIRDFKND IDILVATDVAA  
 Sbjct: 241 LSIVFGRTKRRVDETRGLKLRGFRAEGIHGDLQNKRLRVIRDFKNDQIDILVATDVAA 300

Query: 301 RGLDISGVTHVYNYDIPQDPESYVHRIGRTGRAGKSGSITFVSPNEMGYLTIIENLTKK 360  
 RGLDISGVTHVYNYDI QDPESYVHRIGRTGRAGKSG+SITFVSPNEMGYL++IENLTKK  
 Sbjct: 301 RGLDISGVTHVYNYDITQDPESYVHRIGRTGRAGKSGESITFVSPNEMGYLSMIENLTKK 360

Query: 361 RMTGMKPATASEAFQAKKKVALKRIARDFEQELVSKFDKFKADALELATQYTPEELALY 420  
 +M ++PATA EAFQAKKKVALK+I RDF D+ + S FDKFK DA++LA ++TPEELALY  
 Sbjct: 361 QMKPLRPATAEEAFQAKKKVALKKIERDFADETIRSNFDFKFGDAVQLAAEFTPEELALY 420

Query: 421 VLSLTVQDPESLPEVEITREKPLPFKPSGGGF---KGKGGRG--NGRGGDRRRNRDRGDR- 474  
 +LSLTVQDP+SLPEVEI REKPLPFK GGG GKGGRG N GDRR RGDR  
 Sbjct: 421 ILSLTVQDPDSLPEVEIAREKPLPFKYVGGGHGKNKGKGGGRDRNRNRGDRRGGRGDRN 480

Query: 475 -----RGNRDRDDRGSRCDFKRRDDKFKKDNRRQENKKPHKNTSSEKQTGFVI 522  
 R RD D DFKR+ + KD +E K SS K TGFVI  
 Sbjct: 481 RDERDGRRRQKRDKRDGHGSGNRDFKRSKRNSKDFNKEKK-----SSAKNTGFVI 534

Query: 523 RNKGD 527  
 R+KG+  
 Sbjct: 535 RHKGE 539

A related GBS gene <SEQ ID 8991> and protein <SEQ ID 8992> were also identified. Analysis of this protein sequence reveals the following:

RGD motif 471-473

The protein has homology with the following sequences in the databases:

58.9/74.7% over 494aa

Lactobacillus reuteri

GP|4409804| autoaggregation-mediating protein Insert characterized

ORF01926(301 - 1785 of 2184)

GP|4409804|gb|AAD20136.1|AF091502(1 - 495 of 497) autoaggregation-mediating protein {Lactobacillus reuteri}

%Match = 37.3

%Identity = 58.8 %Similarity = 74.6

Matches = 290 Mismatches = 118 Conservative Sub.s = 78

42 72 102 132 162 192 222 252

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2179

A DNA sequence (GBSx2296) was identified in *S.agalactiae* <SEQ ID 6735> which encodes the amino acid sequence <SEQ ID 6736>. This protein is predicted to be outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

-2459-

Possible site: 19  
>>> May be a lipoprotein

----- Final Results -----

5                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10       >GP:CAB73036 GB:AL139076 putative periplasmic protein [Campylobacter  
          jejuni]  
          Identities = 89/237 (37%), Positives = 132/237 (55%), Gaps = 3/237 (1%)  
  
15       Query: 40   ITVATYSKPTSTFLDLVKDNVKEKGYTLKVVMSDYIQANIALENKEHDANLLQHEFFMS 99  
                  IT+       P + L+L+KD+ K KGY LK+V   SDYI   N ALE KE DANL QH+ F+  
      Sbjct: 23   ITIGATPNPFGSLLELMKDDFKNKGVELKIVEFSYIILPNRALEEKELDANLYQHKPFLE 82  
  
      Query: 100   IFNKENDGHLVSITPIYHSLAGFYGQHLKNIAELKDGAKVAIPSDPANMTRALLLQEKK 159  
                  +N +   +L++ TP+   + G Y + +KN+   LK+GA+VAIP+D   N +RAL LL++ K  
20       Sbjct: 83   EYNLKKGSNLIATTPVLIAPVGVYSKKIKNLENLKEGARVAIPNDATNESRALELLEKAK 142  
  
      Query: 160   LITLKNSTSKKTKAIEDIITNPKKLRIEPPVALNLNQAIFYDLVFNFPGYVTKINLVPKR 219  
                  LI L   + KT   DI   NPKKL+   +   L +A   + D+       +   L P +  
25       Sbjct: 143   LIELNKNTLKTPL--DINKNPKKLKFIELKAAQLPRALDDVDIAITNSNFALGAGLNPSK 200  
  
      Query: 220   DRLLYEKKPDIFAGALVAREDNKNNSDKIKVLKEVLTSEIRHYITKEIPSEAAVAF 276  
                  D +   E K +   +   +V R + KNS+K KV+ E+L S + +   I +       AF  
30       Sbjct: 201   DTIFREDK-NSPYVNYVVVRSEGNSEKTKVIDEILRSDFKAIINEHYKDILIPAF 256

30       SEQ ID 6736 (GBS126) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 7; MW 32kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2180

35       A DNA sequence (GBSx2297) was identified in *S.agalactiae* <SEQ ID 6737> which encodes the amino acid sequence <SEQ ID 6738>. This protein is predicted to be probable permease of ABC transporter. Analysis of this protein sequence reveals the following:

Possible site: 34  
>>> Seems to have no N-terminal signal sequence

40       INTEGRAL   Likelihood = -11.99   Transmembrane   190 - 206 ( 187 - 215)  
          INTEGRAL   Likelihood = -8.44   Transmembrane   25 - 41 ( 16 - 45)  
          INTEGRAL   Likelihood = -6.48   Transmembrane   69 - 85 ( 68 - 90)  
          INTEGRAL   Likelihood = -3.77   Transmembrane   90 - 106 ( 88 - 109)  
45       INTEGRAL   Likelihood = -1.44   Transmembrane   145 - 161 ( 145 - 161)

----- Final Results -----

50                   bacterial membrane --- Certainty=0.5798 (Affirmative) < succ>  
                  bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                  bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55       >GP:AAG08889 GB:AE004963 probable permease of ABC transporter  
          [Pseudomonas aeruginosa]  
          Identities = 80/206 (38%), Positives = 127/206 (60%), Gaps = 4/206 (1%)  
  
      Query: 15   SFWETNMLGLTLILCFLIAFPTGILLFSLRKSZYLIKHSLAYQLNLFLGTLRSVPFLIF 74  
                  +FW   MLG +L+   ++   P G+LLF       + +   Y LL+L +   LRS+PF+I  
      Sbjct: 24   TFW----MLGGSLLFTTVLGLPLGVLLFLTGPQRMFEQKAVYTLSSLVVNILRSLPFIIL 79

-2460-

Query: 75 IFILIPLNRLIFGTSFGTIAAILPLTLVSVSLYARYVEQALLNIPQVVDRALSLGANKR 134  
 + ++IPL LI GTS G AI PL + + +AR VE AL + + +++ ++GA+ R  
 Sbjct: 80 LIVMIPLTVLITGTSGLVAGAIPLLVGATPFFARLVETALREVVDKGIIEATQAMGASTR 139

Query: 135 QIIYYFLIPSIKIDLVLSFTATAISILGYSTIMGVIGAGGLGEYAYRFGYQEYDYPVMYL 194  
 QII+ L+P + ++ + T TAI+++ Y+ + GV+GAGGLG+ A RFGYQ + VM +  
 Sbjct: 140 QIIWNALLPEARPGIIAAITVTAITLVSYTAMAGVVGAGGLGDLAIRFGYQRFQTDVMVV 199

Query: 195 IVVLFITIVFILQSLGYFIANRYSRK 220  
 VV+ +I V ILQ++G + +SRK  
 Sbjct: 200 TVVMLLILVQILQTVGDKLVVHFSRK 225

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2181**

A DNA sequence (GBSx2298) was identified in *S.agalactiae* <SEQ ID 6739> which encodes the amino acid sequence <SEQ ID 6740>. This protein is predicted to be ABC transporter, ATP-binding protein (oppF). Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

|                                                               |
|---------------------------------------------------------------|
| bacterial cytoplasm --- Certainty=0.5454(Affirmative) < succ> |
| bacterial membrane --- Certainty=0.0000(Not Clear) < succ>    |
| bacterial outside --- Certainty=0.0000(Not Clear) < succ>     |

A related GBS nucleic acid sequence <SEQ ID 9333> which encodes amino acid sequence <SEQ ID 9334> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC22280 GB:U32744 ABC transporter, ATP-binding protein  
 [Haemophilus influenzae Rd]  
 Identities = 62/174 (35%), Positives = 104/174 (59%), Gaps = 2/174 (1%)

Query: 1 MKMINGLIPYDKGNIYYQGKEVKSFSDNKLRQMRKDIAYIFQNHNLLAGESVYVYHLALVY 60  
 ++ +N L G++ G E+ SD +L R+ I IFQ+ NLL+ +V+ ++AL  
 Sbjct: 48 IRCVNLLKPTSGSVIVDGVELTKLSDRVLVLARRQIGMIFQHFNLSSRTVFENVALPL 107

Query: 61 KLNHQKVN--HDAINDILDFLGLMDLKQVKCHSLSGGQQQKVAIAMAVLQKPKLILCDEI 118  
 +L + + I +LD +GL + + +LSGGQ+Q+VAIA A+ PK++LCDE  
 Sbjct: 108 ELESSEKAKIQEKITALLDLVGLSEKRDAYPSNLSGGQKQRVATARALASDPKVLCLDEA 167

Query: 119 SSALDTNSEKEIFNLLSDLREKYGISILMIAHHLKQYCDRVMILDHQTIVD 172  
 +SALD + + I LL ++ GI+IL+I H + ++KQ CD+V ++D +V+  
 Sbjct: 168 TSALDPATTQSILKLLKEINRTLGITILLITHEMEVVKQICDQVAVIDQGRIVE 221

There is also homology to SEQ ID 76.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2182**

A DNA sequence (GBSx2299) was identified in *S.agalactiae* <SEQ ID 6741> which encodes the amino acid sequence <SEQ ID 6742>. Analysis of this protein sequence reveals the following:

Possible site: 21

-2461-

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

10   Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2183

A DNA sequence (GBSx2300) was identified in *S.agalactiae* <SEQ ID 6743> which encodes the amino acid sequence <SEQ ID 6744>. Analysis of this protein sequence reveals the following:

15   Possible site: 18  
      >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20                   bacterial cytoplasm --- Certainty=0.0904 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9741> which encodes amino acid sequence <SEQ ID 9742> was also identified.

25   The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB87515 GB:AF034138 unknown [Bacillus subtilis]  
 Identities = 74/125 (59%), Positives = 92/125 (73%)

30   Query: 5   MGIFSGLMGNASQMDTDKVENQLSDILISDEQVDLAYTLIRDLIVFTNYRLILVDKQGVT 64  
           MG   GL+GNAS + T V+ +L+ IL+ E+V+ A+ L+RDLIVFT+ RLILVDKQG+T  
   Sbjct: 1   MGFIDGLLGNASTLSTAAVQEELAHILLEGEKVEAAFKLVRLIVFTDKRLILVDKQGIT 60

35   Query: 65   GKKVSYNSIPYASISRFTVETSGHFDLDAELKIWISSAIEPAEVLQFKNDRNIVSIQKAL 124  
           GKK + SIPY SISRF+VET+G FDLD+ELKIWIS A PA QFK D +I IQK L  
   Sbjct: 61   GKKTEFQSIPIKYSISRFSVETAGRFDLDSSELKIWISGAELPAVSKQFKKDESIYDIQKVL 120

Query: 125 ATAVL 129

A +

40   Sbjct: 121 AAVCM 125

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2184

45   A DNA sequence (GBSx2301) was identified in *S.agalactiae* <SEQ ID 6745> which encodes the amino acid sequence <SEQ ID 6746>. Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

50   ----- Final Results -----

                  bacterial cytoplasm --- Certainty=0.0921 (Affirmative) < succ>

-2462-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9331> which encodes amino acid sequence <SEQ ID 9332> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA74739 GB:Y14370 peptide chain release factor 3  
 [Staphylococcus aureus]  
 Identities = 274/462 (59%), Positives = 349/462 (75%), Gaps = 9/462 (1%)

Query: 1 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAVMVVDSAKGI 60  
 M +E++RGISVTSSVMQFDY +NILDTPGHEDFSEDYRTLMAVD+AVMV+D AKG+  
 Sbjct: 57 MKVEQERGISVTSSVMQFDYDDYEINILDTPGHEDFSEDYRTLMAVDSAVMVIDCAKGV 116

Query: 61 EAQTKKLFVVKHRNIPVFTFINKLDRDGREPLDLEEELEEVGLIASYPMNWPIGMGKSF 120  
 E T KLF+V K R IP+FTFINKLDR G+EP +LL+E+EE L I +YPMNWPIGMG+SF  
 Sbjct: 117 EPPTLKLKFKVCKMRGIPIFTFINKLDRVGKEPFELDEIEETLNIETYPMNWPIGMGQSF 176

Query: 121 EGLYDLHNKRLELYKGDERFASIEDG-----DQLFANNPFYEQVKEDIELLQEAGNDFSE 175  
 G+ D +K +E ++ +E + D D N+ +EQ E++ L++EAG F  
 Sbjct: 177 FGIIDRKSKEIEPRDEENILHLNDDFELEDHAITNDSDFEQAIIEELMLVEEAGEAFDN 236

Query: 176 QAILDGDLTVPVFFGSALTNFGVQTFLDTFLEFAPEPHGHKTTEGNVIDPLAKDFSGFVFK 235  
 A+L GDLTPVFFGSAL NFGVQ FL+ +++FAP P+ +T E + P FSGF+FK  
 Sbjct: 237 DALLSGDLTPVFFGSALANFGVQNFNLNAYVDFAPMPNARQTKENVEVSPFDDSFSGFIFK 296

Query: 236 IQANMDPRHRDRIAFVRIVSGEFERGMGNLRTGKGAKLSNVTQFMAES-RENVTNVA 294  
 IQANMDP+HRDRIAF+R+VSG FER + + L +K S+V + + ++ ++ V +AVA  
 Sbjct: 297 IQANMDPKHRDRIAFMRVVSAGAFER-VWMLLCNVLIKSKRSHVQRHLWQTIKKLVNHAVA 355

Query: 295 GDIIGVYDTGTQYQVGDTLTVGKKNFEFEPLPTFTPELFMKVSAKNVMKQKSFHKGIEQLV 354  
 GDIIG+YDTG YQ+GDTL GK + F+ LP FTPE+FMKVSANVMKQK FHKGIEQLV  
 Sbjct: 356 GDIIGLYDTGNQIGDTLVGGKQTYSFQDLQPFTPEIFMKVSAKNVMKQKHFHKGIEQLV 415

Query: 355 QEGAIQLYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTPMGKKTVRW--INSD 412  
 QEGAIQ YK T + +LGAVGQLQFEV+HRM+ EYN +VVM P+G+K RW N D  
 Sbjct: 416 QEGAIQYYKTLHTNQIILGAVGQLQFEVFEHRMKNEYNVDDVMEPVGRKIARWDIENEDQ 475

Query: 413 LDERMSSSRNIIAKDRFDQPVFLFENDFALRWFADKYPDVKL 454  
 + ++M++SR+IL KDR+D VFLFEN+FA RWF +K+P++KL  
 Sbjct: 476 ITDKMNTSRNIIKDRYDDLVLFLFENEFATRWFEKFPFIKL 517

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6747> which encodes the amino acid sequence <SEQ ID 6748>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2070(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 447/458 (97%), Positives = 455/458 (98%)

Query: 1 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAVMVVDSAKGI 60  
 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAVMVVDSAKGI  
 Sbjct: 57 MDIEKQRGISVTSSVMQFDYAGKRVNILDTPGHEDFSEDYRTLMAVDAAVMVVDSAKGI 116

Query: 61 EAQTKKLFVVKHRNIPVFTFINKLDRDGREPLDLEEELEEVGLIASYPMNWPIGMGKSF 120  
 EAQTKKLFVVKHRNIPVFTFINKLDRDGREPL+LLEEELEEVGLIASYPMNWPIGMG++F  
 Sbjct: 117 EAQTKKLFVVKHRNIPVFTFINKLDRDGREPLELLEEELEEVGLIASYPMNWPIGMGRAF 176



-2463-

Query: 121 EGLYDLHNKRLLEYKGDERFASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD 180  
 EGLYDLHNKRLLEYKGDERFASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD  
 Sbjct: 177 EGLYDLHNKRLLEYKGDERFASIEDGDQLFANNPFYEQVKEDIELLQEAGNDFSEQAILD 236

Query: 181 GDLTPVFFGSALTNFGVQTFDFTFLEFAPEPHGHKTTEGNVIDPLAKDFSGFVFKIQANM 240  
 GDLTPVFFGSALTNFGVQTFDFTFLEFAPEPHGHKTTEGNV+DPLAKDFSGFVFKIQANM  
 Sbjct: 237 GDLTPVFFGSALTNFGVQTFDFTFLEFAPEPHGHKTTEGNVVDPLAKDFSGFVFKIQANM 296

Query: 241 DPRHRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAESRENTNAVAGDIIGV 300  
 DP+HRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAESRENTNAVAGDIIGV  
 Sbjct: 297 DPKHRDRIAFVRIVSGEFERGMGVNLTRTGKGAKLSNVTQFMAESRENTNAVAGDIIGV 356

Query: 301 YDTGTYQVGD TLTVGKKNKFEEPLPTFTPELFMKVSAKNVMKQKSFHKGIEQLVQEGAIQ 360  
 YDTGTYQVGD TLTVGKKNKFEEPLPTFTPE+FMKVS KNVMKQKSFHKGIEQLVQEGAIQ  
 Sbjct: 357 YDTGTYQVGD TLTVGKKNKFEEPLPTFTPEIFMKVSPKNVMKQKSFHKGIEQLVQEGAIQ 416

Query: 361 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRWINSDDLDERMSSS 420  
 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRWI+ DDL+RMSSS  
 Sbjct: 417 LYKNYQTGEYMLGAVGQLQFEVFKHRMEGEYNAEVVMTMGKKTVRWISEDDLQRMSSS 476

Query: 421 RNILAKDRFDQPVFLFENDFALRWFADKYPDVKLEEK 458  
 RNILAKDRFDQPVFLFENDFALRWFADKYPDV LEEKM  
 Sbjct: 477 RNILAKDRFDQPVFLFENDFALRWFADKYPDVTLEEK 514

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2185

A DNA sequence (GBSx2302) was identified in *S.agalactiae* <SEQ ID 6749> which encodes the amino acid sequence <SEQ ID 6750>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3061(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC38046 GB:AF000954 No definition line found [Streptococcus mutans]  
 Identities = 122/142 (85%), Positives = 138/142 (96%)

Query: 1 MLEFAAQKTGKENKEMAVTFVTNERSHELNLLEYRDTDRPTDVISLEYKPEVDISFDEEDL 60  
 +LEFAAQKTGKE+KEMAVTFVTNERSHELNL+YRDT+RPTDVISLEYKPE +SFDEEDL  
 Sbjct: 23 ILEFAAQKTGKEDKEMAVTFVTNERSHELNLKYRDTNRPTDVISLEYKPESLSFDEEDL 82

Query: 61 AENPELAEMLED FDSYIGELFISIDKAKEQAEEYGHSEYEREMGFLAVHGFLHINGYDHYT 120  
 A++P+LAE+L +FD+YIGELFIS+DKA+EQA+EYGHSEYEREMGFLAVHGFLHINGYDHYT  
 Sbjct: 83 ADDPDLAELVLTEDAYIGELFISVDKAREQAEEYGHSEYEREMGFLAVHGFLHINGYDHYT 142

Query: 121 PEEKEMFSLQEEILTAYGLKR 142  
 P+EEKEMFSLQEEIL AYGLKR  
 Sbjct: 143 PQEEKEMFSLQEEILDAYGLKR 164

There is also homology to SEQ ID 120.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-2464-

**Example 2186**

A DNA sequence (GBSx2303) was identified in *S.agalactiae* <SEQ ID 6751> which encodes the amino acid sequence <SEQ ID 6752>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -15.39 | Transmembrane | 108 - 124 ( 100 - 131) |
| INTEGRAL | Likelihood = -8.92  | Transmembrane | 61 - 77 ( 52 - 82)     |
| INTEGRAL | Likelihood = -5.36  | Transmembrane | 41 - 57 ( 40 - 60)     |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.7156(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]

Identities = 107/133 (80%), Positives = 121/133 (90%), Gaps = 2/133 (1%)

Query: 1 MDLNDN--NHKKWKNRNLTSSEFAVTGIFTAFKEERNMRKHLVSAAILVILAGLTFQVSM 58

MDL DN + KKWKNRNLTSSEFAVTGIFTAFKEERNM+KH VSA+L ++AGL F+VS+

Sbjct: 3 MDLRDNKQSQKKWKNRNLTSSEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFKVSV 62

Query: 59 VEWLFLLLSIFLVITFEIINSAIENVVDLASNYHFSMLAKNAKDMAAGAVLVVSLFAVLV 118

+EWLFLLLSIFLVITFEI+NSAIENVVDLAS+YHFSMLAKNAKDMAAGAVLV+S FA L

Sbjct: 63 IEWLFLLLSIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLVISGFAALT 122

Query: 119 GLIIFIPKILALL 131

GLIIF+PKI LL

Sbjct: 123 GLIIFVPKIWFL 135

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6753> which encodes the amino acid sequence <SEQ ID 6754>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -10.67 | Transmembrane | 63 - 79 ( 41 - 84)     |
| INTEGRAL | Likelihood = -7.32  | Transmembrane | 110 - 126 ( 105 - 129) |
| INTEGRAL | Likelihood = -5.41  | Transmembrane | 43 - 59 ( 41 - 62)     |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5267(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the databases:

>GP:AAC38047 GB:AF000954 diacylglycerol kinase [Streptococcus mutans]

Identities = 104/135 (77%), Positives = 119/135 (88%)

Query: 1 MALHDNNTTKRKWKNRNLTSSEFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60

M L DN +++KWKNRNLTSSEFALTG+FTAFKEERN++ H +SA LA +AGL F +S

Sbjct: 3 MDLRDNKQSQKKWKNRNLTSSEFALTGIFTAFKEERNMKKHAVSALLAVIAGLVFKVSV 62

Query: 61 IEWLFLLLAIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLMISGYAVLT 120

IEWLFLLL+IFLVIT EIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVL+ISG+A LT

Sbjct: 63 IEWLFLLLSIFLVITFEIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLVISGFAALT 122

Query: 121 GLIIFIPKIWNIFVH 135

GLIIF+PKIW + H

Sbjct: 123 GLIIFVPKIWFLFH 137

An alignment of the GAS and GBS proteins is shown below.

-2465-

Identities = 98/129 (75%), Positives = 115/129 (88%), Gaps = 2/129 (1%)

Query: 1 MDLNDNN--HKWKNTLTSSMEFAVTGIFTAFKEERNMRKHLVSAILVILAGLTFQVSM 58  
M L+DNN +KWKNT+TSS+EFA+TG+FTAFKEERN+R HL+SA L +AGL F +S

Sbjct: 1 MALHDNNTTKRKWKNTTSSLEFALTGVFTAFKEERNLRSHLLSACLACVAGLFFSISA 60

Query: 59 VEWLFLLLSIFLVITFEIINSAIENVVDLASNYHFSMLAKNAKDMAAGAVLVVSLFAVLV 118  
+EWLFLLL+IFLVIT EI+NSAIENVVDLAS+YHFSMLAKNAKDMAAGAVL++S +AVL

Sbjct: 61 IEWLFLLLAIFLVITLIVNSAIENVVDLASDYHFSMLAKNAKDMAAGAVLMISGYAVLT 120

Query: 119 GLIIFIPKI 127

GLIIFIPKI

Sbjct: 121 GLIIFIPKI 129

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2187

A DNA sequence (GBSx2304) was identified in *S.agalactiae* <SEQ ID 6755> which encodes the amino acid sequence <SEQ ID 6756>. This protein is predicted to be GTPase Era (era). Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1871(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10017> which encodes amino acid sequence <SEQ ID 10018> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD41632 GB:AF072811 GTPase Era [Streptococcus pneumoniae]

Identities = 273/299 (91%), Positives = 290/299 (96%)

Query: 16 MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTETEQIVFIDT 75  
MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTT+ EQIVFIDT

Sbjct: 1 MTFKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTTRNKIMGIYTTDKEQIVFIDT 60

Query: 76 PGIHKPKTALGDFMVESAYSTLREVETVLFMVPADKRGKGDDMIERLKAAPVILVI 135

PGIHKPKTALGDFMVESAYSTLREV+TVLFMVPADKRGKGDDMIERLKAAPVILV+

Sbjct: 61 PGIHKPKTALGDFMVESAYSTLREVDTVLFMVPADKRGKGDDMIERLKAAPVILVV 120

Query: 136 NKIDKVHPDQLLEQIDDFRSQMDFKFVVPISALQGNNVPTLIKLLTDNLEEGFYFPEDQ 195

NKIDKVHPDQLL QIDDFR+QMDFKF+VPISALQGNNV L+ +L++NL+EGFYFP DQ

Sbjct: 121 NKIDKVHPDQLLSQIDDFRNQMDFKFIVPISALQGNVSRVLVDLSENLEEGFYFPSDQ 180

Query: 196 ITDHPERFLVSEMVRKVLHLTQQEVPHSVAVVSVSMKRDEETDKVHIRATIMVERDSQK 255

ITDHPERFLVSEMVRKVLHLT++E+PHSVAVVSVSMKRDEETDKVHIRATIMVERDSQK

Sbjct: 181 ITDHPERFLVSEMVRKVLHLTREIIPHSVAVVSDMKRDEETDKVHIRATIMVERDSQK 240

Query: 256 GIIIGKQGAMLKKIGKMARRDIELMLGDKVYLETVVVKVKNWRDKKLDLADFGYNEKEY 314

GIIIGK GAMLKKIG MARRDIELMLGDKV+LETWVKVKNWRDKKLDLADFGYNE+EY

Sbjct: 241 GIIIGKGGAMLKKIGSMARRDIELMLGDKVFLETWVKVKNWRDKKLDLADFGYNEREY 299

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6757> which encodes the amino acid sequence <SEQ ID 6758>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

-2466-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1088(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 295/297 (99%), Positives = 296/297 (99%)

10 Query: 18 FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTRNKIMGIYTTETEQIVFIDTPG 77  
 FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTRNKIMGIYTTETEQIVFIDTPG  
 Sbjct: 2 FKSGFVAILGRPNVGKSTFLNHVMGQKIAIMSDKAQTRNKIMGIYTTETEQIVFIDTPG 61

15 Query: 78 IHKPKTALGDFMVESAYSTLREVETVLFMVPADKRGKGDDMIERLKAAKIPVILVINK 137  
 IHKPKTALGDFMVESAYSTLREVETVLFMVPADKRGKGDDMIERLKAAKIPVILVINK  
 Sbjct: 62 IHKPKTALGDFMVESAYSTLREVETVLFMVPADKRGKGDDMIERLKAAKIPVILVINK 121

20 Query: 138 IDKVHPDQLEQIDDFRSQMDFKFVVPISALQGNNVPTLIKLLTDNLEEGFQYFPEDQIT 197  
 IDKVHPDQLEQIDDF SQMDFKFVVPISAL+GNNVPTLIKLLTDNLEEGFQYFPEDQIT  
 Sbjct: 122 IDKVHPDQLEQIDDFHSQMDFKFVVPISALEGNNVPTLIKLLTDNLEEGFQYFPEDQIT 181

25 Query: 198 DHPERFLVSEMVRKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQKGI 257  
 DHPERFLVSEMVRKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQKGI  
 Sbjct: 182 DHPERFLVSEMVRKVLHLTQQEVPHSVAVVVESMKRDEETDKVHIRATIMVERDSQKGI 241

30 Query: 258 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY 314  
 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY  
 Sbjct: 242 IIGKQGAMLKKIGKMARRDIELMLGDKVYLETWVKVKKNWRDKKLDLADFGYNEKEY 298

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 2188**

A DNA sequence (GBSx2305) was identified in *S.agalactiae* <SEQ ID 6759> which encodes the amino acid sequence <SEQ ID 6760>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2679(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 2189**

A DNA sequence (GBSx2306) was identified in *S.agalactiae* <SEQ ID 6761> which encodes the amino acid sequence <SEQ ID 6762>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have a cleavable N-term signal seq.

## ----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

-2467-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAA16793 GB:D90900 hypothetical protein [Synechocystis sp.]  
 Identities = 36/119 (30%), Positives = 57/119 (47%), Gaps = 15/119 (12%)

Query: 390 TSDYEKAKVIHDHLVNNYTYATEELATTRETASGISIHAEALYKDKRGVCQAFVAMFKD 449  
 ++D+E+A++ + + N Y +A TR I PE + +C ++ ++  
 10 Sbjct: 153 SNDWEERLAYSWITQNIAYDVP-MAETRN----IDDLRPETVLARGETICSGYSNLYQA 207

Query: 450 MAATAGLSVWYVTGQAGGG-----NHAWNIVTINGVKYVDTTWDNNIKSNKYF 498  
 +A GL V + G A GG NHAWN V I+G Y +DTTW I S+ F  
 15 Sbjct: 208 LAKELGLDVVIIIEGFAKGGDIVGDDPDVNHAWNGVKIDGQWYLLDTTWGAGIVSDGKF 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6763> which encodes the amino acid sequence <SEQ ID 6764>. Analysis of this protein sequence reveals the following:

Possible site: 23

20 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 41/181 (22%), Positives = 79/181 (42%), Gaps = 17/181 (9%)

30 Query: 355 ITITYTLKGDMLHKEYKQFVDSFVKENITNKNITSDEYKAKVIHDHLVNNYTYATE-- 412  
 + +T+ + D +++++ Q + + + N +K+ YE+ K ++ ++ + Y +  
 Sbjct: 124 VFVTFPIPEDAKNIYQDL-QAIGNDIVANTPSKD---RYEQVKYFYEVIIIRD TDYNKKAF 179

35 Query: 413 ELATTRETASGISIHAEALYKDKRGVCQAFVAMFKDMAATAGLSVWYVTGQAGGGN--- 469  
 E + A S +++++ D VC +A F+ + AG+ V Y+ G  
 Sbjct: 180 EAYQSGSQAQVASNQDIKSVFIDHLSVCNGYAQAFQFLCQKAGIPVAYIRGTGTSQQPQQ 239

Query: 470 ---HAWNIVTINGVKYVDTTW-----DNNIKSNKYFLVGKTIMDADHLLDSQYNALAKDI 522  
 HAWN V IN Y VD TW DN++ K + + + L + + +KDI  
 40 Sbjct: 240 SFAHAWNVAQINNTYYGVDVTWGDVFDNHLSHQKQGTINYSFLCLPDYLMALSHQPSKDI 300

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 45 Example 2190

A DNA sequence (GBSx2307) was identified in *S.agalactiae* <SEQ ID 6765> which encodes the amino acid sequence <SEQ ID 6766>. This protein is predicted to be rgg protein. Analysis of this protein sequence reveals the following:

Possible site: 29

50 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.16 Transmembrane 187 - 203 ( 187 - 203)

----- Final Results -----

bacterial membrane --- Certainty=0.1065(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2468-

A related GBS nucleic acid sequence <SEQ ID 10015> which encodes amino acid sequence <SEQ ID 10016> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAA26968 GB:M89776 rgg [Streptococcus gordonii]
    Identities = 71/273 (26%), Positives = 140/273 (51%), Gaps = 16/273 (5%)

    Query: 8  KELGKTLRRLRKGGKVSISSLADEHLSKSQISRFERGESEITCSRLNLNLDKLNITIDEF 67
              K  GK L+ +R+ K +S+ +A  +S +Q+SR+ERG S +T  + L  +++++ EF
    Sbjct: 5  KSSGKILKIRESKNMSLKEVAAGDISVAQLSRYERGISLTVDSFYSLRNMSVSLAEF 64

10  Query: 68  VSI-HSKAHTFFILLNRVRKYCAEKNVTKLVALL-----EDHNHKDYEKIMIK 115
              + H+      +L  ++ +  E N+ KL ++L      E  N+K  I+I+
    Sbjct: 65  QVYVHNRYREADDVVLSQKLSEAQRENNIVKLESILAGSEAMAQEFPEKKNYK-LNTIVIR 123

15  Query: 116 ALIFSIDQSIEPNQEELARLTDYLFVEQWGYEIIILLGNC SRLINYNLFLLTKE MVNS 175
              A + S +  + ++ ++  LTDYLF+VE+WG YE+ L N  L+  TL  EM+N
    Sbjct: 124 ATLTSNPDYQVSKGDIEFLTDYLFVVEEWGRYELWLFNSVNLTLTLETTFASEMINR 183

20  Query: 176 FAYSEQNKTNKILVTQLAINCLIIISIDHSYFEHSHYLIDKVRSLLDQEVNFYEKTVFLYV 235
              +      N+  + ++ +N +  I++++ + +  ++ + +  E + Y++ +  Y
    Sbjct: 184 TQFYNNLPENRRRIIKMLLNVSACIENNLQVAMKFLNYIDNTKIPETDLYDRVLIKYH 243

    Query: 236 TGYHHLKLGDTSSGKEDMRKALQIFKYLGEDSF 268
              Y  K+G+  + + D+ + L  F+YL  DSF
25  Sbjct: 244 KALYSYKVGNPFA-RHDIEQCLSTFEYL--DSF 273
  
```

There is also homology to SEQ ID 628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 30 Example 2191

A DNA sequence (GBSx2308) was identified in *S. agalactiae* <SEQ ID 6767> which encodes the amino acid sequence <SEQ ID 6768>. Analysis of this protein sequence reveals the following:

```

    Possible site: 36
    >>> Seems to have no N-terminal signal sequence

35  ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3234(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40
  
```

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:BAA05066 GB:D26071 formamidopyrimidine-DNA glycosylase
      [Streptococcus mutans]
    Identities = 182/271 (67%), Positives = 217/271 (79%)

45  Query: 1  MPPELPEVETVRKGLERLVNQEIASITIKVPMVKTDLNDFMISLPGKTIQQVLRRGKYL 60
              MPPELPEVETVR+GLE L+V ++I S+ ++VPMVKT + DF + + G+T + + RRGKYL
    Sbjct: 1  MPPELPEVETVRGLEHLIVGKKIVSVEVRVPMVKTVGVEDFQLDILGQTFESIGRRGKYL 60

50  Query: 61  LFDGEMVMVSHLRMEGKYLLFPNKVPDNKHFHLYFKLTNGSTLVYQDVRKFGTFELVRK 120
              L +      ++SHLRMEGKYLLF ++VPDNKHFHL+F L  GSTLVYQDVRKFGTFEL+ K
    Sbjct: 61  LLNLNRQTIISHLRMEGKYLLFEDEVDPDNKHFHLFFGLDGGSTLVYQDVRKFGTFELLPK 120

55  Query: 121 SSLKDYFTQKKLGPEPTADTFQFEPFSKGLANSKKPIKPLLLDQRLVAGLGNIVDEV LW 180
              S ++ YF QKK+GPEP A  F+ +PF +GLA S K IK LLLDQ LVAGLGNIVDEV LW
    Sbjct: 121 SQVEAYFVQKKIGPEPNAKDFKLKPFEEGLAKSHKVIKTLLEDQHLVAGLGNIVDEV LW 180

    Query: 181 AAKIHPQRLANQLTESETSLHKEIIRILTLGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240
  
```

-2469-

AAK+ P+RLA+QL SE +H E IRIL L IEKGGSTIR+YKN+LGEDG+MQ LQVY  
 Sbjct: 181 AAKVDPERLASQLKTSEIKRIHDETIRILQLAIEKGGSTIRSYKNSLGEDGSMQDCLQVY 240

Query: 241 GKTGQPCPRCGCLIKKIKVGGRGTHYCPRCQ 271  
 GKT QPC RC I+KIKVGGRGTH+CP CQ  
 Sbjct: 241 GKTDPQPCARCATPIEKIKVGGRGTHFCPCSQ 271

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6769> which encodes the amino acid sequence <SEQ ID 6770>. Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2068(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 190/271 (70%), Positives = 229/271 (84%)  
 Query: 1 MPPELPEVETVRKGLERLVVNQEIASITIKVPMVKTDLNDFMISLPGKTIQQVLRRGKYL 60  
 MPPELPEVETVR+GLE LV+ QEI ++T+KVPKMVKTDL F ++LPG+ IQ V RRGKYL  
 Sbjct: 1 MPPELPEVETVRRGLETLVLGQEIVAVTLKVPKMVKTDLETFTALTPGQIIQSVGRRGKYL 60  
 Query: 61 LFDGEMVMVSHLRMEGKYLLFPNKVPDNKHFHLYFKLTNGSTLVYQDVRKFGTFELVRK 120  
 L D G++V+VSHLRMEGKYLLFP++VPDNKHFH++F+L NGSTLVYQDVRKFGTF+L+ K  
 Sbjct: 61 LIDLQQLVLVSHLRMEGKYLLFPDEVPDNKHFHVFFELKNGSTLVYQDVRKFGTFDLIAK 120  
 Query: 121 SSLKDYFTQKKGPEPTADTFQFEPFSKGLANSKKPIKPLLLDQRLVAGLGNIVDEVILW 180  
 S L +F ++KLGPEP +TF+ + F L +S+KPIKP LLDQ LVAGLGNIVDEVILW  
 Sbjct: 121 SQLSAFFAKRKLGPPEPKKETFKLKTFEALLSSQKPIKPHLLDQTLVAGLGNIVDEVILW 180  
 Query: 181 AAKIHPQRIANQLTSETSLHKEIIRILTLGIEKGGSTIRTYKNALGEDGTMQKYLQVY 240  
 AAK+HP+ +++L ++E LH E IRIL LGIEKGGST+RTY+NALG DGTMQ YLQVY  
 Sbjct: 181 AAKVHPETASSRLNKAIEIKRLHDETIRILALGIEKGGSTVRYRNALGADGTMQDYLQVY 240  
 Query: 241 GKTGQPCPRCGCLIKKIKVGGRGTHYCPRCQ 271  
 G+TG+PCPRCG I K+KVGGRGTH CP+CQ  
 Sbjct: 241 GQTKGPCPRCGQAIVKLKVGGRGTHICPKCQ 271

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2192

A DNA sequence (GBSx2309) was identified in *S.galactiae* <SEQ ID 6771> which encodes the amino acid sequence <SEQ ID 6772>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0797(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10013> which encodes amino acid sequence <SEQ ID 10014> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00353 GB:AF008220 YtaG [Bacillus subtilis]

-2470-

Identities = 80/189 (42%), Positives = 113/189 (59%), Gaps = 1/189 (0%)

Query: 8 MTKIIGLTGGIASGKSTVTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGP EILD 67  
 MT +IGLTGGIASGKSTV ++ E G VIDAD + + KG Y+ +++ G +IL  
 Sbjct: 1 MTLVIGLTGGIASGKSTVANMLIEKGITVIDADIIAKQAVEKGMPAYRQIIDEFGEDILL 60

Query: 68 ADGELDRPKLSQMIFANPDNMKTSARLQNSIIRQELACQDQLKQTEEIF-FMDIPLLE 126  
 ++G++DR KL ++F N + + +RQE+ +RD+ E F +DIPL E  
 Sbjct: 61 SNGDIDRKLKLGALVFTNEQKRLALNAIVHPAVRQEMLNRRDEAVANREAFVVLIDIPLLFE 120

Query: 127 EKYIKWFDEIWLVFVDKEKQLQRLMARNNYSREEAELRLSHQMPLTDKKS FASLIIDNNG 186  
 K D+I +V V KE QL+RLM RN + EEA R+ QMPL +K + A +IDN+G  
 Sbjct: 121 SKLESLVDKIIIVSVTKELQLERLMKRNQLTEEEAVSRIRSQMPLEEK TARADQVIDNSG 180

Query: 187 DLITLKEQI 195  
 L K Q+  
 Sbjct: 181 TLEETKRQL 189

A related sequence was also identified in GAS <SEQ ID 9111> which encodes the amino acid sequence  
 <SEQ ID 9112>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 59  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty= 0.101(Affirmative) < succ>  
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>  
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 118/191 (61%), Positives = 153/191 (79%)

Query: 9 TKIIGLTGGIASGKSTVTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGP EILDA 68  
 T IIG+TGGIASGKSTV K+IR++G++VIDADQVVH LQ KGG+LY+AL E G +IL A  
 Sbjct: 9 TMIIGITGGIASGKSTVVKVIRKAGYQVIDADQVVHDLQEKGGRLYEALREAFGNQILKA 68

Query: 69 DGELDRPKLSQMIFANPDNMKTSARLQNSIIRQELACQDQLKQTEEIFFMDIPLLEEK 128  
 DGELDR KLS+M+F+NPDM TS+ +QN II++ELA +RD L Q++ IFFMDIPLL+E  
 Sbjct: 69 DGELDRPKLSEMLFNPDMMATSSAIQNQIIKEELAARDHLAQSQAIFFMDIPLLMELG 128

Query: 129 YIKWFDEIWLVFVDKEKQLQRLMARNNYSREEAELRLSHQMPLTDKKS FASLIIDNNGDL 188  
 Y WFD IWL+VD + QLQRLMARN + +A R++ Q+P+ +KK +ASL+IDN+GD+  
 Sbjct: 129 YQDWFDAIWLVIYVDAQTLQRLMARNRLDKGARQRIASQLPIEEKPYASLVIDNSGDI 188

Query: 189 ITLKEQILDAL 199  
 L +Q+ AL  
 Sbjct: 189 AALIKQVQSAL 199

A related GBS gene <SEQ ID 8993> and protein <SEQ ID 8994> were also identified. Analysis of this  
 protein sequence reveals a signal peptide at residues 1-16.

The protein has homology with the following sequences in the databases:

42.2/60.6% over 189aa

OMNI|NT01BS3382| Insert characterized

ORF02237(319 - 885 of 1206)  
 OMNI|NT01BS3382(3 - 192 of 200) ()  
 %Match = 17.0  
 %Identity = 42.1 %Similarity = 60.5  
 Matches = 80 Mismatches = 74 Conservative Sub.s = 35

78 108 138 168 198 228 258 288  
 KNSPTAFG\*SIDRI\*NLKITQGNYSHFNFHRKRWLHD\*NI\*ECSWRGRYDAKVFTGLW\*NWATVSKVWLFN\*EDKSRRE



5  
RDALLPSVSMMLTKIIGLTGGIASGKSTVTKI IRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGP EILDADAGELDRPK  
| : | : ||||| ||||| : : | | |||| : : || | : : : : | : | : : || |  
VDLLTLVIGLTGGIASGKSTVANMLEIKGITVIDADI IAKQAVEKGMPAYRQI IDEFGEDILLSNGDIDRKK  
10 20 30 40 50 60 70

10  
LSQMIFANPDNMKTSARLQNSIIRQELACQRDLKQTETEEIFF-MDIPLLEEKYIKWFD EIWLVFVDKEKQLQRLMARNN  
| : : | | : : : ||| : : | | : |||| | | : | : | | | | : ||| | |  
LGALVFTNEQKRRLALNAIVHPAVRQEMLNRRDEAVANREAFVVL D I PLLFESKLES L VDKI IVSV T K E L Q L E R L M K R N Q  
90 100 110 120 130 140 150

15  
YSREEAELRLSHQMPLTDKKS FASLIIDNNGDLITLKEQIL DALQRL\*NY\*MDNVFIHFLSL LH \*F\*KTC D\*T TVIVQ\*Y  
: ||| | : |||| : | : | : ||| : | | | : : :  
LTEEEAVSRIRSOMPLEEKTARADQVIDNSGTLEETKRQLDEIMNSWA  
170 180 190 200

Score = 218 bits (550), Expect = 4e-59  
Identities = 104/175 (59%), Positives = 138/175 (78%)

```

25 Query: 25 VVKVIRKAGYQVIDADQVVHDIQEKGGRLYEALREAFGNQILKADGELDRTKLSEMLFSN 84
      V K+IR++G++VIDADQVVH LQ KGG+LY+AL E G +IL ADGELDR KLS+M+F+N
      Sbjct: 20 VTKIIRESGFKVIDADQVVHKLQAKGGKLYQALLEWLGPEILDADGELDRPKLSQMIFAN 79

30 Query: 85 PDNMATSSAIQNQIIKEELAAKRDHLAQSQAIFFMDIPLLMEELGYQDWFDAIWL VYVDAQ 144
      PDNM TS+ +QN II++ELA +RD L Q++ IFFMDIPLL+E Y WFD IWL V+VD +
      Sbjct: 80 PDNMKTSARLQNSIIRQELACQRDQLKQTEIIFMDIPLLIEEKYIKWFDEIWL VFVDKE 139

      Query: 145 TQLQRLMARNRLDKGARQRIASQLPIEEKKPYASLVIDNSGDI AALIKQVQSAL 199
      QLQRLMARN + +A R++ Q+P+ +KK +ASL+IDN+GD+ L +Q+ AL
35 Sbjct: 140 KQLQRLMARNNYSREEAELRLSHOMPLTDKKSFASLIIDNNGDLITLKEQILDAL 194

```

40 The purified GST fusion product was used to immunise mice and the resulting antiserum was used for FACS (Figure 278). This confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 A DNA sequence (GBSx2310) was identified in *S.agalactiae* <SEQ ID 6775> which encodes the amino acid sequence <SEO ID 6776>. Analysis of this protein sequence reveals the following:

```

50      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4073(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA30330 GB:AP000005 253aa long hypothetical ATP-binding

-2472-

transport protein [Pyrococcus horikoshii]  
 Identities = 78/240 (32%), Positives = 130/240 (53%), Gaps = 13/240 (5%)

Query: 3 LVIRDIRKRFQETEVLRGASRYFYSKITGVLRNGAGKTTLFNILEYGDLAADNGTICLL 62  
 +++ ++RK+F EVL+G ++ G+I G+LG NG+GK+T IL G + G + +  
 Sbjct: 2 IIVENLRKKFGSKEVLKGINFTVNDGEIYGLLGPNGSGKSTTMRILSGIITDFEGKVMVA 61

Query: 63 -KDNHEYPLTDKDI-GIVYSENYLPEFLTGYEFVKFYMDLH--PSDDL-MTIDDYLD FME 117  
 D P+ K+I G V L E LT EF F + P D L + +D  
 Sbjct: 62 GVDVSRDPMKVKEIVGYPETPALYESLTPAEFFSFIGGVRRIPQDILEERVKRLVDAFG 121

Query: 118 IGQTERHRIIKGYSDGMKSKLSLICLMISKPKVILLDEPLTAVDVSSIAIKRLLELSE 177  
 IG+ +++I S G K K+SLI ++ P+V++LDE + +D S+ + LL E E  
 Sbjct: 122 IGK-YMNQLIGTLSFGTKQKISLISALLHDPQVLILDEAMNGLDPKSARIFRELLFEFKE 180

Query: 178 D-HIIILSTHIMALAEDLCDIVAVLDKQKL---QTLDIDR---KHEQFEERLLQVLKGDE 230  
 + I+ STHI+ALAE +CD + ++ +G++ T+D R + E+ E+ L++ + E  
 Sbjct: 181 EGKSIVFSTHILALAEVMDRIGIIEGRIVAEGTIDELREIAREEKLEDIFLKLTAKE 240

20 There is also homology to SEQ ID 2876.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2194

25 A DNA sequence (GBSx2311) was identified in *S.agalactiae* <SEQ ID 6777> which encodes the amino acid sequence <SEQ ID 6778>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.6138(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 2195

40 A DNA sequence (GBSx2312) was identified in *S.agalactiae* <SEQ ID 6779> which encodes the amino acid sequence <SEQ ID 6780>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

45 INTEGRAL Likelihood = -15.34 Transmembrane 526 - 542 ( 511 - 546)  
 INTEGRAL Likelihood = -9.61 Transmembrane 340 - 356 ( 335 - 359)  
 INTEGRAL Likelihood = -8.17 Transmembrane 455 - 471 ( 451 - 476)  
 INTEGRAL Likelihood = -8.01 Transmembrane 97 - 113 ( 95 - 121)  
 INTEGRAL Likelihood = -8.01 Transmembrane 216 - 232 ( 207 - 236)  
 INTEGRAL Likelihood = -3.40 Transmembrane 50 - 66 ( 46 - 67)  
 INTEGRAL Likelihood = -1.33 Transmembrane 178 - 194 ( 178 - 194)

50 ----- Final Results -----  
 bacterial membrane --- Certainty=0.7135(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2473-

A related GBS nucleic acid sequence <SEQ ID 10011> which encodes amino acid sequence <SEQ ID 10012> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is  
5 homology to SEQ ID 376.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2196

10 A DNA sequence (GBSx2314) was identified in *S.agalactiae* <SEQ ID 6781> which encodes the amino acid sequence <SEQ ID 6782>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have no N-terminal signal sequence

15  
INTEGRAL Likelihood = -8.17 Transmembrane 140 - 156 ( 134 - 160)  
INTEGRAL Likelihood = -6.64 Transmembrane 255 - 271 ( 253 - 274)  
INTEGRAL Likelihood = -5.79 Transmembrane 345 - 361 ( 343 - 363)  
INTEGRAL Likelihood = -3.29 Transmembrane 184 - 200 ( 183 - 202)  
INTEGRAL Likelihood = -2.34 Transmembrane 66 - 82 ( 65 - 83)  
INTEGRAL Likelihood = -1.65 Transmembrane 221 - 237 ( 221 - 239)  
20 INTEGRAL Likelihood = -0.00 Transmembrane 121 - 137 ( 121 - 137)

----- Final Results -----

25 bacterial membrane --- Certainty=0.4270(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9401> which encodes amino acid sequence <SEQ ID 9402> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump  
[Streptococcus pneumoniae]  
Identities = 213/372 (57%), Positives = 295/372 (79%)

35 Query: 1 MPFMVLYVEQLGAPSNKVEWYAGLSVLSALSSALVAPLWGRADKYGRKPMVMVRAGLMM 60  
+PFM ++VE LG S +V +YAGL++S+SA+S+AL +P+WG LADKYGRKPM+RAGL M  
Sbjct: 28 VPFMPPIFVENLGVGSQQVAFYAGLAISVSAISAALFSPIWGILADKYGRKPMIRAGLAM 87

40 Query: 61 TFTMGGLAFIHSVTGLLILRLNGIFAGYVPNSTALIASQAPQEEESGYALGTLATGVTGG 120  
T TMGGLAF+ ++ L+ LR+LNG+FAG+VPN+TALIASQ P+E+SG ALGTL+TGV G  
Sbjct: 88 TITMGGLAFVPNIYWLIFLRLNGVFPAGFVPNATALIASQVPKEKSGSALGTLSTGTVVAG 147

45 Query: 121 MLIGPLLGGLLAEWFGIREVFLLVGTILLISTLMTIFMVKEDFKPISNEETMPTTEVFKS 180  
L GP +GG +AE FGIR VFLLVG+ L ++ ++TI +KEDF+P++ E+ +PT E+F S  
Sbjct: 148 TLTGPFIFGFIAELEFGIRTVFLLVGSFLFLAAILTICFIKEDFQPVAKKEKAIPTELFTS 207

50 Query: 181 VKSLQILIGLGFVTSMTIIQISAQSIAPILTYIRHLGQTENLMFVSGLIVSGMGFSSILSS 240  
VK +L+ LF+TS +IQ SAQSI PIL LY+R LGQTENL+FVSGLIVS MGFSS++S+  
Sbjct: 208 VKYPYLLLNLFLLTSFVIQFSAQSIGPILALYVRDLGQTENLLFVSGLIVSSMGFSSMMSA 267

55 Query: 241 PKLGRIGDRIGNHRLLLLALLYSFLMYVLCSLAQTSLQLGVIRFLYGFGTGALMPSINSI 300  
+G++GD++GNHRL++A YS ++Y+LC+ A + LQLG+ RFL+G GTGAL+P +N++  
Sbjct: 268 GVMGKLGDKVGNHRLLVVAQFYSVIYLLCANASSPLQLGLYRFLFGLGTGALIPGVNAL 327

Query: 301 LTKIAPRQGLSRIFSYNQMFSNLQGVLPFVGSVSIHLGFRWVFFVTSFIVLANFVWCF 360  
L+K+ P+ G+SR+F++NQ+F LG V+GP GSAV+ G+ VF+ TS V + ++

-2474-

Sbjct: 328 LSKMTPKAGISRVAFAFNQVFFYLGGVVGPMAGSAVAGQFGYHAVFYATSLCVAFSCLFNL 387

Query: 361 INFRKYIRVKEI 372

I FR ++VKEI

5 Sbjct: 388 IQFRTLKVKKEI 399

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6783> which encodes the amino acid sequence <SEQ ID 6784>. Analysis of this protein sequence reveals the following:

Possible site: 58

10 &gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -10.14 Transmembrane 165 - 181 ( 150 - 185)

INTEGRAL Likelihood = -7.43 Transmembrane 371 - 387 ( 367 - 391)

INTEGRAL Likelihood = -3.88 Transmembrane 90 - 106 ( 86 - 109)

INTEGRAL Likelihood = -3.35 Transmembrane 145 - 161 ( 143 - 162)

15 INTEGRAL Likelihood = -1.70 Transmembrane 279 - 295 ( 279 - 297)

INTEGRAL Likelihood = -0.85 Transmembrane 209 - 225 ( 209 - 226)

INTEGRAL Likelihood = -0.27 Transmembrane 347 - 363 ( 347 - 363)

----- Final Results -----

20 bacterial membrane --- Certainty=0.5055(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the databases:

25 &gt;GP:CAA07482 GB:AJ007367 multi-drug resistance efflux pump

[Streptococcus pneumoniae]

Identities = 236/396 (59%), Positives = 309/396 (77%)

30 Query: 1 VNWQRNLKVAWLGNFFTGASFSLVMPFMALYVENLGTPTLVEYYAGLAVAVTALASALF 60

+NW+ NL++AW GNF TGAS SLV+PFM ++VENLG ++ V +YAGLA++V+A+++ALF

Sbjct: 4 INWKDNLRIAWFGNFLTGTASISLVVPFMPFIFVENLGVGSQQVAFYAGLAISVSAISALF 63

Query: 61 APVWGKLADRYGRKPMMLRASFMVFTMGGLAIIPNVFWLLILRLLTGVSAGYVPNATAL 120

+P+WG LAD+YGRKPM+RA MT TMGGLA +PN++WL+ LRL L GV AG+VPNATAL

35 Sbjct: 64 SPIWGILADKYGRKPMIRAGLAMTITMGGLAFVNIYWLIFLRLNGVFAGFVPNATAL 123

Query: 121 IASQAPKEESGYALGTLATGVTAGALIGPLLGGILAEELGIRQVFLLVGVILFLCSLMTA 180

IASQ PKE+SG ALGTL+TGV AG L GP +GG +AEL GIR VFLLVG LFL +++T

Sbjct: 124 IASQVPKEKSGSALGTLSTGVVAGTLTGPFIFGFIAELEFGIRTVFLLVGSFLFLAAILTI 183

40 Query: 181 VYVKEEFKPVRRFEMIPTKVILKQVKSPQIMLGLFVTSMIQISAQSVAPILSLYIRHLG 240

++KE+F+PV + + IPTK + VK P ++L LF+TS +IQ SAQS+ PIL+LY+R LG

Sbjct: 184 CFIKEDFQVAKKAIPTKELFTSVKYPYLLNLFLTSTFVIQFSAQSIGPILALYVRDLG 243

45 Query: 241 QTHNLMFTSGLVVSAMGFSSSLFSSSYLGKLGDRGNHRLLLAALCYSFIMYFSSALAQTS 300

QT NL+F SGL+VS+MGFSS+ S+ +GKLG+ GNHRL+ A YS I+Y A A +

Sbjct: 244 QTENLLFVSGSLIVSSMGFSSMSAGVMGKLGDKVGNHRLLVQAQFYSVIYLLCANASSP 303

50 Query: 301 FQLGLVLRFAFGFGALMPSINSLLTKLTPKEGISRVFAYNQMFNSNLQVIGPFIGSNVA 360

QLG+ RF +G G GAL+P +N+LL+K+TPK GISRVFA+NQ+F LG V+GP GS VA

Sbjct: 304 LQLGLYRFLFGLGTGALIPGVNALLSKMTPKAGISRVAFAFNQVFFYLGGVVGPMAGSAVA 363

Query: 361 VVLGYRSVFYVTSLVFVNLIWSLIIFRKYIKVKDI 396

GY +VFY TSL V + +++LI FR +KVK+I

55 Sbjct: 364 GQFGYHAVFYATSLCVAFSCLFNLIQFRTLKVKKEI 399

An alignment of the GAS and GBS proteins is shown below.

Identities = 262/373 (70%), Positives = 314/373 (83%)

60 Query: 1 MPFMVLYVEQLGAPSNNKVEWYAGLSVLSALSSALVAPLWGRIADKYGRKPMVVRAGLMM 60

MPFM LYVE LG P+ VE+YAGL+V+++AL+SAL AP+WG+LAD+YGRKPM+RA +M

Sbjct: 25 MPFMALYVENLGTPTLVEYYAGLAVAVTALASALFAPVWGKLADRYGRKPMMLRASFM 84

Query: 61 TFTMGGLAFIHSVTGLLILRLINGIFAGYVPNSTALIASQAPQEESGYALGTLATGVTGG 120

-2475-

TFTMGGLA I +V LLILR+L G+ AGYVPN+TALIASQAP+EESGYALGTLATGVT G  
 Sbjct: 85 TFTMGGLAIIPNVFWLLILRLLTGVSAGYVPNATALIASQAPKEESGYALGTLATGVTAG 144  
 Query: 121 MLIGPLLGLLAEWFGIREVFLLVGTILLISTLMTIFMVKEDFKPISNEETMPTTEVFKS 180  
 5 LIGPLLGG+LAE GIR+VFLLVG IL + +LMT VKE+FKP+ E +PT + K  
 Sbjct: 145 ALIGPLLGGILAEELGIRQVFLLVGVILFLCSLMTAVYVKEEFKPVRRFEMIPTKVILKQ 204  
 Query: 181 VKSLQILIGLFVTSMI IQISAQSIAPILTYIRHLGQTENLMFVSGLIVSGMGFSSILSS 240  
 10 VKS. QI++GLFVTSMI IQISAQS+APIL+LYIRHLGQT NLMF SGL+VS MGFSS+ SS  
 Sbjct: 205 VKSPQIMLGLFVTSMI IQISAQSVAPILSLYIRHLGQTHNLMFTSGLVVSAMGFSSLFSS 264  
 Query: 241 PKLGRIGDRIGNHRLLLALLYSFLMYVLCSLAQTSLQLGVIRFLYGFGTGALMPSINSI 300  
 LG++GDR GNHRLLL AL YSF+MY +LAQTS QLGV+RF YGFG GALMPSINS+  
 15 Sbjct: 265 SYLGKLGDRFGNHRLLLAALCYSFIMYFSSALAQTSLQLGVLRFA YGFGV GFGV GFGV GALMPSINSI 324  
 Query: 301 LTKIAPRQGLSRIFSYNQMFSNLGQVLGPFVGS AVSIHLGFRWVFFVTSFIVLANFVWCF 360  
 LTK+ P++G+SR+F+YNQMFSNLGQV+GPF+GS V++ LG+R VF+VTS IV N +W  
 Sbjct: 325 LTKLTPKEGISRVFAYNQMFSNLGQVIGPFIGSNVAVVLGYRSVFYVTSFIVFNLIWSL 384  
 20 Query: 361 INFRKYIRVKEIV 373  
 I FRKYI+VK+IV  
 Sbjct: 385 IIFRKYIKVKDIV 397

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 25 vaccines or diagnostics.

### Example 2197

A DNA sequence (GBSx2315) was identified in *S.agalactiae* <SEQ ID 6785> which encodes the amino  
 acid sequence <SEQ ID 6786>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 30 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2343(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB69986 GB:U94356 glycerol kinase [Enterococcus faecalis]  
 Identities = 156/186 (83%), Positives = 167/186 (88%), Gaps = 1/186 (0%)  
 40 Query: 3 SEEKYIMAI DQGTSSRAIIFNKKGEK IASSQKEFPQIFPQAGWVEHNANQIWNSVQSVI 62  
 +EEKYIMAI DQGTSSRAIIF+KKG KI SSQKEF Q FP AGWVEHNAN+IWNSVQSVI  
 Sbjct: 2 AEEKYIMAI DQGTSSRAIIFDKKGNKIGSSQKEFTQYFPNAGWVEHNANEIWNVSVQSVI 61  
 45 Query: 63 AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122  
 AG+ IES +KP I IGITNQRETTVVWDK TGLPIYNAIVWQSRQT PIADQLK++G+  
 Sbjct: 62 AGSLIESGVKPTDIAGIGITNQRETTVVWDKATGLPIYNAIVWQSRQTTPIADQLKEDGY 121  
 50 Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPQAQERAEGELLFGTIDTTLVWVWKLTDGLVHV 182  
 + MIHEKTGL+IDAYFSATKVRWILDHV GAQERAE GEL+FGTIDTTLVWVWKL T G HV  
 Sbjct: 122 SEMIHEKTGLI IDAYFSATKVRWILDHVEGAQERAENGELMFGTIDTTLVWVWKLTDGLVHV 180  
 Query: 183 TDYSNA 188  
 TDYSNA  
 55 Sbjct: 181 TDYSNA 186

There is also high homology to SEQ ID 2844:

Identities = 174/186 (93%), Positives = 182/186 (97%)  
 60 Query: 3 SEEKYIMAI DQGTSSRAIIFNKKGEK IASSQKEFPQIFPQAGWVEHNANQIWNSVQSVI 62  
 S+EKYIMAI DQGTSSRAIIFN+KGEK++SSQKEFPQIFP AGWVEHNANQIWNSVQSVI

-2476-

Sbjct: 2 SQEKYIMAIQDQTTSSRAIIFNQKGEKVSSSQKEFPQIFPHAGWVEHNANQIWNSVQSVI 61

Query: 63 AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122  
AGAFIESSIKP QIEAIGITNQRETTVVWDKKTG+PIYNAIVWQSRQTAPIA+QLKQ+GH

5 Sbjct: 62 AGAFIESSIKPSQIEAIGITNQRETTVVWDKKTGVPIYNAIVWQSRQTAPIAEQLKQDGH 121

Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGLVHV 182  
T MIHEKTGLVIDAYFSATK+RWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDG VHV

10 Sbjct: 122 TKMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEKGELLFGTIDTWLVWKLTDGAVHV 181

Query: 183 TDYSNA 188  
TDYSNA

Sbjct: 182 TDYSNA 187

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2198

A DNA sequence (GBSx2317) was identified in *S.agalactiae* <SEQ ID 6787> which encodes the amino acid sequence <SEQ ID 6788>. This protein is predicted to be glycyl-tRNA synthetase beta chain (glyS).

- 20 Analysis of this protein sequence reveals the following:

Possible site: 18  
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2933(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 30 >GP:CAB14468 GB:Z99117 glycyl-tRNA synthetase (beta subunit)  
[Bacillus subtilis]  
Identities = 315/687 (45%), Positives = 447/687 (64%), Gaps = 21/687 (3%)
- 35 Query: 3 KDLLLELGLLELPAYVVTPESEKQLGQKMKVFLDHRLSFETVQIFSTPRRLAVRVKGLAD 62  
+DLLLE+GLEE+PA + S QLG K+ +L++ ++ V++F+TPRRLAV VK +A+  
Sbjct: 4 QDLLLEIGLEEMPARFLNESMVQLGDKLTGWLKEKNITHGEVKLFNTPRRLAVFVKDVAE 63
- 40 Query: 63 QQTDLTDFKGPSKKIALDAEGNFSKAAQGFVRGKGLSVDDIEFREVKGEEYVYVTKHET 122  
+Q D+ E+ KGP+KKIALDA+GN++KAA GF +G+G +V+D+ +EVKG EYV+V K +  
Sbjct: 64 KQDDIKEEAKGPAKKIALDADGNWTKAAIGFSKGQGANVEDLYIKEVKGIEYVVFVQKFQA 123
- 45 Query: 123 GKSAIDVLASVTEVLTELTFPVMNHWANNSFEYIRPVHTLVVLLDDQALELDFLDIHSGR 182  
G+ +L ++ ++T L FP NM W N YIRP+ +V L + ++ SGR  
Sbjct: 124 GQETKSLLEPELSGLITSLHFPKNMRWGNEDLRYIRPIKWIVALFGQDVIPFSITNVESGR 183
- 50 Query: 183 ISRGHRFLGSDTEISSASSYEDDLRQQFVIADAKERQQMIVNQIHAIEEKKNISVEIDED 242  
++GHRFLG + I S S+YE+ L+ Q VIAD R+QMI +Q+ + + N S+ +DED  
Sbjct: 184 TTQGHRLGHEVSIESPAYEEQLKGQHVADPSVRKQMIQSQLETMAAENNWSIPVDED 243
- 55 Query: 243 LLNEVLNLVEYPTAFLGSFDEKYLDVPEEVLVTSMKNHQRYFVVRDRDGKLLPNFISVRN 302  
LL+EV +LVEYPTA GSF+ ++L +PEEVLVT+MK HQRYF V+D++G LLP+FI+VRN  
Sbjct: 244 LLDEVNHLVEYPTALYGSFESEFLSIPEEVLVTMKEHQRYFPVKDKNGDLLPHFITVRN 303
- 60 Query: 303 GNAEHIENVIKGNEKVLVARLEDEFFWQEDQKLNIALDLVEKLGQVTFHEKIGSLYEHMD 362  
GN+ IENV +GNEKVL ARL D FF++EDQKLNI V+KL+ + FHE++GSL + +  
Sbjct: 304 GNSHAIENVARGNEKVLRLRLSDASFFYKEDQKLNIDANVKKLENIVFHEELGSLADKVR 363
- Query: 363 RVKVISQYLAEKADLSDEEKLAVLRAASIYKFDLLTGMVDEFDELQGIMGEKYALLAGEQ 422  
RV I++ LA + ++ V RAA I KFDL+T M+ EF ELQGIMGEKYA + GE
- Sbjct: 364 RVTSIAEKLAVRLQAEDTLKHVKRAAEISKFDLVTHMIYEFPELQGIMGEKYARMLGED 423

-2477-

Query: 423 PAVAAAIREHYMPTADGELPETRVGAILALADKFDTLSSFFSVGLIPSGSNDPYALRR 482  
 AVAAA+ EHYMP SA GE P T GA++A+ADK DT+ SFFS+G+IP+GS DPY L R  
 Sbjct: 424 EAVAAVNEHYMPRSAGGETPSTFTGAVVAMADKLDTIASFFSIGVIPTGSQDPYGLPRQ 483

5 Query: 483 TQGIVRILEAFGWDIPLDELVTNLYGLSFASLDYANQKEVMAFISARIEKMIGS-KVPKD 541  
 GIV IL W I +EL+T F D N E++ F + R++ ++ + ++ D  
 Sbjct: 484 ASGIVAILLDRNWGISFEELLT-----FVQTDKEN--ELLDFFTQRLKYVLNAEQIRHD 535

10 Query: 542 IREAVLESPTYIVSLILEASQALVQKSKDAQYKVSVESLSRAFNLAEKVTHSVLVDSSLF 601  
 + +AVLES L +Q L QK +K + E+L R +++K + LF  
 Sbjct: 536 VIDAVLESSELEPYSAHLKAQVLEQKLGAPGFKETAELGRVISISKKGVRGD-IQPDLF 594

15 Query: 602 ENNQEKALYQAILSLLELTEDMHDNLDK-----LFALSPIINDFFDNTMVMTDDEKM 652  
 EN E L+ A + + E++ +N K L AL I+ +FD+TMV+ D+E +  
 Sbjct: 595 ENEYEAKLFDAYQTAK--ENLQENFSKKDYEALASLAALKEPIDAYFDHTMVIADNESL 652

20 Query: 653 KQNRLAILNSLVAKARTVAAFNLLNTK 679  
 K NRLA + SL + ++ A N L K  
 Sbjct: 653 KANRLAQMVSLADEIKSFANMNALIVK 679

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2835> which encodes the amino acid sequence <SEQ ID 2836>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.96 Transmembrane 450 - 466 ( 450 - 466)

----- Final Results -----

bacterial membrane --- Certainty=0.1383(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 505/679 (74%), Positives = 578/679 (84%)

35 Query: 1 MTKDLLLELGLLELPAYVVTTPSEKQLGQKMKVFLDHRLSFETVQIFSTPRRLAVRVKGL 60  
 M+K+LL+ELGLEELPAYVVTTPSEKQLG+++ FL ++RLSFE +Q FSTPRRLAVRV GL  
 Sbjct: 1 MSKNLLIELGLEELPAYVVTTPSEKQLGERLATFLTENRLSFEDIQTFTSTPRRLAVRVSG 60

40 Query: 61 ADQQTDLTDFKGP+KKIALDA+GNFSKAAQGFVRGKGLSVDDIEFREVKGEEYVYVTKH 120  
 ADQQTDLTDFKGP+KKIALDA+GNFSKAAQGFVRGKGL+ D IEFREVKGEEYVYVTKH  
 Sbjct: 61 ADQQTDLTDFKGP+KKIALDA+GNFSKAAQGFVRGKGLTTDAIEFREVKGEEYVYVTKH 120

45 Query: 121 ETGKSAIDVLASVTEVLTETLTFPVNMHWANNSFEYIRPVHTLVLLDDQALELDFLDIHS 180  
 E GK A +VL VTEVL+ +TFPV+MHWANNSFEYIRPVHTL VLL+D+ALELDFLDIHS  
 Sbjct: 121 EAGKPAKEVLLGVTEVLSAMTFPVSMHWANNSFEYIRPVHTLVLLNDEALELDFLDIHS 180

50 Query: 181 GRISRGHRFLGSDTEISSASSYEDDLRQQFVIADAKERQQMIVNQIHAIEKKNISVEID 240  
 GR+SRGHRFLG++T I+SA SYE DLR QFVIADAKERQ+MIV QI +E ++ + V+ID  
 Sbjct: 181 GRVSRGHRFLGTETTTITSADSYEADLRSQFVIADAKERQEMIVEQIKTLEVEQGVQVDID 240

55 Query: 241 EDLLNEVLNLVEYPTAFSGFDEKYLDPPEEVLVTSMKNHQRYFVVRDRDGKLLPNFISV 300  
 EDLLNEVLNLVE+PTAF+GSF+ KYLDVPEEVLVTSMKNHQRYFVVRD+ G L+PNF+SV  
 Sbjct: 241 EDLLNEVLNLVEFPTAFMGSFYAKYLDPPEEVLVTSMKNHQRYFVVRDQAGHLPNFVSV 300

60 Query: 301 RGNNAEHIENVIKNEKVLVARLEDGEFFWQEDQKLNIAIDLVEKLRQVTFHEKIGSLYEH 360  
 RGN + IENVIKNEKVLVARLEDGEFFW+EDQKL IADLV KL VTFHEKIGSL EH  
 Sbjct: 301 RGNNDQAIENVIKNEKVLVARLEDGEFFWREDQKLQIADLVAKLTNVTTFHEKIGSLAEH 360

65 Query: 361 MDRVKVISQYLAEKADLSDEEKLAVLRAASIYKFDLLTGMVDFDELQGIMGEKYALLAG 420  
 MDR +VI+ LA++A+LS EE AV RAA IYKFDLLTGMV EFDELQGIMGEKYALLAG  
 Sbjct: 361 MDRTRVIAASLAKEANLSAEVTAVDRAAQIYKFDLLTGMVGEFDELQGIMGEKYALLAG 420

Query: 421 EQPAVAAAIREHYMPTADGELPETRVGAILALADKFDTLSSFFSVGLIPSGSNDPYALR 480  
 E AVA AIREHY+P +A G LPET+VGA+LALA K DTLSSFFSVGLIPSGSNDPYALR  
 Sbjct: 421 EDAAVATAIREHYLPDAAGGALPETKVGAVLALAAKLDTLSSFFSVGLIPSGSNDPYALR 480

-2478-

Query: 481 RATQGIVRILEAFGWDIPLDELVTNLYGLSFASLDYANQKEVMAFISARIEKMIGSKVPK 540  
 RATQGIVRIL+ FGW IP+D+LV +LY LSF SL YAN+ +VM FI AR++KM+G PK  
 Sbjct: 481 RATQGIVRILDHFGWRIPMDKLVDLSYDLSDLSLTANKADVMNFIRARVDKMMGKAAPK 540

Query: 541 DIREAVLESPTYIVSLILEASQALVQKSKDAQYKVSVESLSRAFNLAEKVTHSVLVDSSL 600  
 DIREA+L S T++V +L A++ALV+ S YK +VESLSRAFNLAEK SV VD SL  
 Sbjct: 541 DIREAILASSTFVPEMLAAAEALVKASHTENYKPAVESLSRAFNLAEKADASVQVDPSSL 600

Query: 601 FENNOEKALYQAILSLLELTEDMHDNLDKLFALSPIINDDNTMVMTDDEKMKQNRLAIL 660  
 FEN QE L+ AI L L L+++FALSP+INDDNTMVM D+ +K NRLAIL  
 Sbjct: 601 FENEQENTLFAAIQGLTLGSAQQLEQVFALSPVINDNTMVMAGDQALKNNRLAIL 660

Query: 661 NSLVAKARTVAAFNLLNTK 679  
 + LV+KA+T+ AFN LNTK  
 Sbjct: 661 SDLVSKAKTIVAFNQLNTK 679

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2199

A DNA sequence (GBSx2318) was identified in *S.agalactiae* <SEQ ID 6789> which encodes the amino acid sequence <SEQ ID 6790>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2182(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD24436 GB:AF112858 NAD(P)H dehydrogenase [Bacillus  
 stearothermophilus]  
 Identities = 64/174 (36%), Positives = 98/174 (55%), Gaps = 6/174 (3%)

Query: 2 NTLIVNSHPDFSNPYSFTTILQEKFIELYNEHFPNHLQSLNLYDCVLPEITKEVLLSIW 61  
 N L + +HP + S++ + + FI+ Y + P+H++ L+LY +PEI +V S W  
 Sbjct: 3 NVLYITAPH-DDTQSYSMVAGKAFIDTYKQVHPDHEVIHLDLKEYIPEIDVDVF-SGW 60

Query: 62 SKQRKGL--ELTADEIVQAKISKDLLEQFKSHHRIVFVSPMHNYNVTARAKTYIDNIFI 118  
 K R G EL+ +E + +L EQF S + VFV+PM N++ K YID + +  
 Sbjct: 61 GKLRSGKSFEELSDEEKAKVGRMNELCEQFISADKYVFVTPMWNFSFPPVLKAYIDAVAV 120

Query: 119 AGETFKYKTENGSVGLMTDDYRLMLLESAGSIYSKGQYSPYEFVHYLKAIFKDF 172  
 AG+TFKYTE G VGL+TD + L +++ G YS+G + E YL I + F  
 Sbjct: 121 AGKTFKYTEQGPVGLLTDK-KALHIQARGGFYSEGPAAEMEMGHRYLSVIMQFF 173

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2200

A DNA sequence (GBSx2319) was identified in *S.agalactiae* <SEQ ID 6791> which encodes the amino acid sequence <SEQ ID 6792>. This protein is predicted to be glycyl-tRNA synthetase (glyQ). Analysis of this protein sequence reveals the following:

Possible site: 56



-2479-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.1364(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9521> which encodes amino acid sequence <SEQ ID 9522> was also identified.

10   The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05089 GB:AP001511 glycyl-tRNA synthetase (alpha subunit)  
                  [Bacillus halodurans]  
   Identities = 222/287 (77%), Positives = 250/287 (86%)

15   Query: 6   LTFQEIIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPSRRPA 65  
               +   Q +ILTLQ++W+ Q C+L+QAYD EKGAGTMSPYT LR IGPEPWNAAYVEPSRRPA  
   Sbjct: 1   MNVQTMILTLQEQYWSKQNCILLQAYDTEKGAGTMSPYTMLRTIGPEPWNVAYVEPSRRPA 60

20   Query: 66   DGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLKSLELLGINPLEHDIRFVEDNWNENPST 125  
               DGRYGENPNRLYQHHQFQV+MKPSP+NIQELYL SL LGINPLEHDIRFVEDNWNENPS  
   Sbjct: 61   DGRYGENPNRLYQHHQFQVIMKPSPTNIQELYLDSLRLGINPLEHDIRFVEDNWNENPSL 120

25   Query: 126   GSAGLGWEVWLDGMEITQFTYFQQVGGGLQTGPVTSEVTYGLERLASYSIQEVDVSVYDIEWA 185  
               G AGLGWEVWLDGMEITQFTYFQQVGGGL+ PV++E+TYGLERLASYSIQ+ ++V+D+EW  
   Sbjct: 121   GCAGLGWEVWLDGMEITQFTYFQQVGGGLEANPVSAEITYGLERLASYSIQDKENVFDLEWV 180

30   Query: 186   PGVKYGEIFTQPEYEHKYSFEISDQVMLLENFEKFEREAKRALEEGLVHPAYDYVLKCS 245  
               G YG+IFTQPEYEHKYS+FE+SD ML E F +E+EA RALEE LV PAYDYVLKCS  
   Sbjct: 181   EGFTYGDIFTQPEYEHKYSKTFEVSDSAMLFELFSTYEKEADRALEENLVFPAYDYVLKCS 240

35   Query: 246   HTFNLLDARGAVSVTERAGYIARIRNLARVVAKTFVAERKKLGFPPL 292  
               HTFNLLDARGA+SVTER GYI R+RNLAR AK + ER+KLGFP+L  
   Sbjct: 241   HTFNLLDARGAISVTERTGYIGRVNRLARKCAKKYEEEREKLGFPML 287

35   A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6793> which encodes the amino acid sequence <SEQ ID 6794>. Analysis of this protein sequence reveals the following:

Possible site: 55

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

40   ----- Final Results -----

            bacterial cytoplasm --- Certainty=0.2081(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45   An alignment of the GAS and GBS proteins is shown below.

Identities = 290/304 (95%), Positives = 294/304 (96%)

50   Query: 2   MSKKLTFQEIIILTLQQFWNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS 61  
               MSKKLTFQEIIILTLQQ+WNDQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS  
   Sbjct: 1   MSKKLTFQEIIILTLQQYWNQGCMLMQAYDNEKGAGTMSPYTFLRAIGPEPWNAAYVEPS 60

55   Query: 62   RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLKSLELLGINPLEHDIRFVEDNWE 121  
               RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYL SLE LGINPLEHDIRFVEDNWE  
   Sbjct: 61   RRPADGRYGENPNRLYQHHQFQVVMKPSPSNIQELYLASLEKLGINPLEHDIRFVEDNWE 120

60   Query: 122   NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGGLQTGPVTSEVTYGLERLASYSIQEVDVSVYD 181  
               NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGGL T PVT+EVTYGLERLASYSIQEVDVSVYD  
   Sbjct: 121   NPSTGSAGLGWEVWLDGMEITQFTYFQQVGGGLATSPVTAEVTYGLERLASYSIQEVDVSVYD 180

60   Query: 182   IEWAPGVKYGIEIFTQPEYEHKYSFEISDQVMLLENFEKFEREAKRALEEGLVHPAYDYV 241  
               IEWAPGVKYGIEIF QPEYEHKYSFEISDQ MLLENFEKFE+EA RALEEGLVHPAYDYV  
   Sbjct: 181   IEWAPGVKYGIEIFLQPEYEHKYSFEISDQMLLENFEKFEKASRALEEGLVHPAYDYV 240

-2480-

Query: 242 LKCSHTFNLLDARGAVSVTERAGYIARIRNLRVVAKTFFAERKKLGFPLLDDEETRIKLL 301  
 LKCSHTFNLLDARGAVSVTERAGYIARIRNLRVVAKTFFAERKKLGFPLLDDE TR LL  
 Sbjct: 241 LKCSHTFNLLDARGAVSVTERAGYIARIRNLRVVAKTFFAERKKLGFPLLDDEATRAILL 300

Query: 302 AEED 305  
 AE+D  
 Sbjct: 301 AEDD 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2201

A DNA sequence (GBSx2320) was identified in *S.agalactiae* <SEQ ID 6795> which encodes the amino acid sequence <SEQ ID 6796>. This protein is predicted to be vacB protein (vacB). Analysis of this protein sequence reveals the following:

Possible site: 60  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2966(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9399> which encodes amino acid sequence <SEQ ID 9400> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15366 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 338/780 (43%), Positives = 485/780 (61%), Gaps = 47/780 (6%)

Query: 4 AKAFPKLIKTISNLESHRQL---RFDDNGSLSLQKKEAKKKEITVRGLFRANKAGFGFL- 59  
 A+ F +L+K + LE + R D G +K ++G A+ GF FL  
 Sbjct: 36 AEEFKELVKALVALEEKGLIVRTRSDRYG-----IPEKMNLIKGISAHAKGFAFL 87

Query: 60 SIDQDEDDMFIGKNDIAYAIIDGDTVEAVVKKPADRLNGTAAEARVVNIVERSLKTIVGKF 119  
 D D+FI N++ A++GD V + + +G+ E V+ I+ER+++ +VG +  
 Sbjct: 88 PEDTSLSDVFIPPNELNTAMNGDIVMRLNSQS---SGSRQEGTVIRILERAIQVRVGT 144

Query: 120 VLDDERPKYAGYIKSNQKINQKIYIRKEPV--VLDGTEIIKVIDIDKYPTRGHDFYFASV 177  
 + G++ ++KI I+I K +G +++ V + YP G V  
 Sbjct: 145 T----ETRNFGFVIPDDKKITSDFIPKNGKNGAAEGHKVV-VKLTSYP-EGRMNABGEV 198

Query: 178 RDIVGHQGDVGIDVLELVESMDIVSEFPEDVIAEANAIPDAPTEKDLIGRVDLRQEVTF 237  
 I+GH+ D GID+L V+ + EFP D + +A++ PD EKDL R DLR +V T  
 Sbjct: 199 ETILGHKNPDGIDILSVIHKHGLPGEFPADAMEQASSTPDTIDEKDLKDRDLRDQVIVT 258

Query: 238 IDGADAKDLDDAVHIKLLDNGHFELGVHIADVSYYVTEGSALNREALSRGTSVYVTD 297  
 IDGADAKDLDDAV + LD+G ++LGVHIADVS+YVTE S +++EAL RGTSVY+ DRV+  
 Sbjct: 259 IDGADAKDLDDAVTVTKLDGSGYKLGVHIADVSHYVTENSPIDKEALERGTSVYLVDRVI 318

Query: 298 PMLPERLSNGICSLNPDLRLTQSCIMEIDQNGRVVNHQITQSVINTYRMTYTAVNDII 357  
 PM+P RLSNGICSLNP +DRLT SC M I+ G+V H+I QSVI TT RMTY+ VN I+  
 Sbjct: 319 PMIPHRLSNGICSLNPKVDRRLTLSCEMTINSQGVTEHEIFQSVIKTTERMTYSVDVNKIL 378

Query: 358 A-GDEEICSEYESIVSSVQHMTLHHTLEAMRTRRGALNFDTSEAKIMVNDKGMFVDIVI 416  
 DEE+ +YE +V + M L L R RGA++FD EAK++V+D+G D+VI  
 Sbjct: 379 VDDDEELKQKYEPPLVPMFKDMERLAQILRDKRMDRGAVDFDFKAEKVLVDDEGAVKDVVI 438

Query: 417 RNRGIAERMIESFMLAANETVAEHYARLKLPTFYRIHEEPKAEKLQKFIDYASVFGVQIQ 476  
 R R +AE+++IE FML ANETVAEH+ + PFIYRIHEEP AEKLQKF+++ + FG ++

-2481-

Sbjct: 439 RERSVAEKLIEEFMLVANETVAEHFHMNVPIYRIHEEPNAEKLQKFLEFVTTFGYVVK 498

Query: 477 GTATKITQSALQDFMKVQGPQSEVLSSMLLRSMQQARYSEHNHGHYGLAAEYTHFTS 536  
GTA I ALQ + V+ +P . V+S ++LRSM+QA+Y + GH+GL+ E+YTHFTS

5 Sbjct: 499 GTAGNIHPRALQSILDAVRDRPEETVISTVMLRSMQAKYDPQSLGHFGLSTEFYTHFTS 558

Query: 537 PIRRYPDLLVHRMIRDY-DDKAMDKA--DHFANLPEIATQTSSLEERRAIDAERIVEAMK 593  
PIRYPDL+VHR+IR Y + +D+A + +A +P+IA TSS+ERRA+DAER + +K

10 Sbjct: 559 PIRYPDLIVHRLIRTYLINGKVEATQEKWAERLPDIAEHTSSMERRAVIDAERETDDLK 618

Query: 594 KAEYMEEYVGEEFEGVVASVVKFGMFVELPNTIEGLIHVTTL-PEYYHFNERTLTLQGEK 652  
KAEYM + +GEEF+G+++SV FGMFVELPNTIEGL+HV+ + +YY F+E+ + GE+

Sbjct: 619 KAEYMLDKIGEEFDGMISSVTNFGMFVELPNTIEGLVHVSFMTDDYYRFDEQHFAMIGER 678

15 Query: 653 SGKVFVRVGQIKVKLIRSDKETGIDIDFDYLPDSDFDIVEKVS KSSREGPNRSSKREHQHR 712  
+G VFR+G +I VK++ +K+ +IDF+ + +G P R + +

Sbjct: 679 TGNVFRIGDEITVKVVDVKNKDERNIDFEIV-----GMKGTPRRPRELD---- 721

20 Query: 713 ISDRDNKNKNTSKKKASRKPKRNSDSKSHHKKDDRTTGSTKKKTKKPFYKGVAKKGQKRK 772  
S R K ++K+ S + S K + T KKK K+ F +K +K+K

Sbjct: 722 -SSRSRKRKGPARKRVQSTNTFVSPAPS-EEKGEWFTKPKKKKKKRGFQNAKPKQKRKKKK 779

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6797> which encodes the amino acid sequence <SEQ ID 6798>. Analysis of this protein sequence reveals the following:

25 Possible site: 30  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0811(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 579/773 (74%), Positives = 664/773 (84%), Gaps = 22/773 (2%)

35 Query: 1 MAGAKAFPKLIKTISNLESHRQLRFDNGSLSLQKKEAKKKEITVRGLFRANKAGFGFLS 60  
MAGAK FP LIKTIS +ES LRF D+GSL+L+K+ KKKE TV+G+FRANKAGFGFL

Sbjct: 27 MAGAKHFP SLIKTISKMESQSLRFSDDGSLALRKEREKKKEPTVQGVFRANKAGFGFLH 86

40 Query: 61 IDQDEDDMFIGKNDIAYIDGDTVEAVVKKPADRLNGTAAEARVNVNIVERSLKT LVGKFV 120  
+D++EDDMFIG+ND+ Y AIDGDTVE VVKKPADRL GTAAEA+VV IV+RSLKT VG F+

Sbjct: 87 VDENEDDMFIGRNDVGYAIDGDTVEVVKKPADRLKGTAAEAKVVAIVDRSLKTAVGTFI 146

45 Query: 121 LDDERPKYAGYIKSKNQKINQKIYIRKEPVVLDGTEI IKVDIDKYPTRGHDFVVASVRDI 180  
LDD++PKYAGYI+SKNQKI QKIYI+KEPVVL GTEI IKVDIDKY PTRGHDFVVASVRDI

Sbjct: 147 LDDDKPKYAGYIRSKNQKIQQKIYIKKEPVVLKGTEI IKVDIDKYPIRGHDFVVASVRDI 206

Query: 181 VGHQGDVGIDVLELVLESMDIVSEFPEDVIAEANAIPDAPTEKDLIGRVDLRQEVTFITIDG 240  
VGHQGDVGIDVLELVLESMDIVSEFP +V+AEANAI +APT KDLIGRVDLRQE T TIDG

50 Sbjct: 207 VGHQGDVGIDVLELVLESMDIVSEFP AEVLA EANAISEAPTAKDLIGRVDLRQETTITIDG 266

Query: 241 ADAKDLDDAVHIKLLDNGHFELGVHIAADVSYVTEGSALNREALSRGTSVYVTRDRVPM 300  
ADAKDLDDA+HIKLLDNG++ELGVHIAADVSYVTEGSAL++EA++RGTSVYVTRDRVPM

Sbjct: 267 ADAKDLDDAIHIKLLDNGNYELGVHIAADVSYVTEGSALDKAEIARGTSVYVTRDRVPM 326

55 Query: 301 PERLSNGICSLNPNLDRLTQSCIMEIDQNGRVVNHQITQSVINTTYRMTYTAVNDI IAGD 360  
PERLSNGICSLNPN+DRLTQS +MEI+ G VVN+QI QSVI TTYRMTY+ VND+IAGD

Sbjct: 327 PERLSNGICSLNPNIDRLTQSCALMEINSQGHVVNYQICQSVIKTTYRMTYSTVNDMIAGD 386

60 Query: 361 EEICSEYESIVSSVQHMVTLHHTLEAMRTRRGALNFDTSEAKIMVNDKGM PVDIVIRNRG 420  
EE E+ SI V MV LH LEAMR++RGALNFDT EAKI+VNDKGM PVD+V+R RG

Sbjct: 387 EEALQEFASIADDVTLMVALHRILEAMRSKR GALNFDTQEAKIIVNDKGM PVDVVL RQRG 446

65 Query: 421 IAERMIESFMLAANETVAEHYARLKL PFIYRIHEEPKAEKLQKFIDYASVFGVQIQGTAT 480  
IAERMIESFMLAANE VAEH+A+ KLPFIYRIHEEPKAEKLQ+ FIDYAS FG+ IQGTA

-2482-

Sbjct: 447 IAERMIESFMLAANECVAEHFAKAKLPFIYRIHEEPKAEKLQQFIDYASTFGIHIQGTAN 506

Query: 481 KITQSALQDFMKVQGPQSEVLMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR 540  
 KI+Q ALQ FM KV+GQPG+EVL+MMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR

5 Sbjct: 507 KISQEALQAFMAKVEGQPGAELNMMLLRSMQQARYSEHNHGHYGLAAEYYTHFTSPIRR 566

Query: 541 YPDLLVHRMIRDYDDKAMDKADHFANLIPEIATQTSSLERRAIDAERIVEAMKKAEMYEE 600  
 YPDLLVHRM+R+Y+ + +K DHFA +IPE+AT +S LERRAIDAER+VEAMKKAEYM E

10 Sbjct: 567 YPDLLVHRMVREYNQPSQEKRDHFAQIPELATSSSQLERRAIDAERVVEAMKKAEMYMAE 626

Query: 601 YVGEEFEGVVASVVKFGMFVELPNTIEGLIHVTTLPEYYHFNERTLTQLQGEKSGKVFRVG 660  
 YVGEEF+G+V+SVVKFG FVELPNTIEGL+H+T+LPEYYHFNERTL+LQGEKSGKVF+VG

Sbjct: 627 YVGEEFDGIVSSVVKFGFFVELPNTIEGLVHITSLPEYYHFNERTLSLQGEKSGKVFKVG 686

15 Query: 661 QQIKVKLIRSDKETGDIDFDYLPSTDFDIVEKVS KSSREGRPNRSSKREHQHRISDRDNKN 720  
 Q I+VKL+++DKETGDIDF+YLPSTDF+VEK+ S + R +R K+

Sbjct: 687 QPIRVKLVKADKETGDIDFEYLPSTDFVVEKIKMSDKASRRDR-----RKS 732

Query: 721 KNTSKKKASRKPKRNSDSKSHHHKDDRTTGSTKKKTKKPFYKGVAKKGQKRKS 773  
 +SK ++PK + +K T G TTK +KKPFYK AKK +++S

20 Sbjct: 733 SKSSKGTKKKEPKEVAKAK-----TKGKTGKSGKPFYKEQAKKSRKRS 777

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 2202

A DNA sequence (GBSx2321) was identified in *S.agalactiae* <SEQ ID 6799> which encodes the amino acid sequence <SEQ ID 6800>. This protein is predicted to be VacB homolog (smpB). Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2988(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC23745 GB:AF052209 VacB homolog [Streptococcus pneumoniae]  
 Identities = 121/155 (78%), Positives = 139/155 (89%)

40 Query: 1 MVKGQGNVVAQNKAHHDYTTIVETIEAGIVLTGTEIKSVRAARITLKDGYAQIKNGEAWL 60  
 M KG+G VVAQNKA HDYTIV+T+EAG+VLTGTEIKSVRAARI LKDG+AQ+KNGE WL

Sbjct: 1 MAKGEGKVVAQNKKARHDYTTIVDTLEAGMVLGTGTEIKSVRAARINLKDGFAQVKNGEVWL 60

45 Query: 61 INVHITPYDQGNINWQDPDRTRKLLKREIEKISNELKGTGMTLVPLKVYLKDGFAKVL 120  
 NVHI PY++GNIWNQ+P+R RKLLL K++I+K+ E KGTGMTLVPLKVY+KDG+AK+L

Sbjct: 61 SNVHIAPYEEGNIWNQEPERRRKLKHKQIQKLEQETKGTGMTLVPLKVYIKDGYAKLL 120

Query: 121 LGLAKGKHHDYDKRESIKRREQNRDIARQLKNYSR 155  
 LGLAKGKHHDYDKRESIKRREQNRDIAR +K N R

50 Sbjct: 121 LGLAKGKHHDYDKRESIKRREQNRDIARVMKAVNQR 155

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6801> which encodes the amino acid sequence <SEQ ID 6802>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2918(Affirmative) < succ>

-2483-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5       Identities = 124/155 (80%), Positives = 145/155 (93%)

Query: 1    MVKQGQNVVAQNKAHHDYITIVETIEAGIVLTGTEIKSVRAARITLKDGYAQIKNGEAWL 60  
           M KG+G+++AQNKKA HDY IVET+EAGIVLTGTEIKSVRAARI LKDG+AQIKNGEAWL  
 10   Sbjct: 1   MAKGECHILAQNKARHDYHIVETVEAGIVLTGTEIKSVRAARIQLKDGFAQIKNGEAWL 60

Query: 61   INVHITPYDQGNINWQDPDRTRKLLKREIEKISNELKGTGMTLVPLKVYLKDGFAKVL 120  
           +NVHI P++QGNINW DP+RTRKLLKREI ++NELKG+GMTLVPLKVYLKDGFAKVL  
 15   Sbjct: 61   VNVHIAPFEQGNINWADPERTRKLLKREITHLANELKSGMTLVPLKVYLKDGFAKVL 120

Query: 121   LGLAKGKHHDYDKRESIKRREQNRDIARQLKNYNSR 155  
           +GLAKGKH+YDKRE+IKRR+Q RDI +Q+K+YN+R  
 20   Sbjct: 121   IGLAKGKHEYDKRETIKRRDQERDIKKQMKHYNAR 155

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 20   vaccines or diagnostics.

### Example 2203

A DNA sequence (GBSx2322) was identified in *S.agalactiae* <SEQ ID 6803> which encodes the amino acid sequence <SEQ ID 6804>. Analysis of this protein sequence reveals the following:

25       Possible site: 14  
       >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.6876(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 30       bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 35   vaccines or diagnostics.

### Example 2204

A DNA sequence (GBSx2323) was identified in *S.agalactiae* <SEQ ID 6805> which encodes the amino acid sequence <SEQ ID 6806>. This protein is predicted to be d-serine/d-alanine/glycine transporter (cycA). Analysis of this protein sequence reveals the following:

40       Possible site: 55  
       >>> Seems to have a cleavable N-term signal seq.

|               |                    |               |                        |
|---------------|--------------------|---------------|------------------------|
| INTEGRAL      | Likelihood = -9.02 | Transmembrane | 71 - 87 ( 62 - 90)     |
| INTEGRAL      | Likelihood = -8.92 | Transmembrane | 320 - 336 ( 316 - 344) |
| INTEGRAL      | Likelihood = -8.33 | Transmembrane | 254 - 270 ( 251 - 275) |
| 45   INTEGRAL | Likelihood = -6.00 | Transmembrane | 158 - 174 ( 154 - 175) |
| INTEGRAL      | Likelihood = -2.76 | Transmembrane | 197 - 213 ( 196 - 213) |
| INTEGRAL      | Likelihood = -2.50 | Transmembrane | 117 - 133 ( 116 - 136) |
| INTEGRAL      | Likelihood = -1.38 | Transmembrane | 282 - 298 ( 279 - 298) |
| 50   INTEGRAL | Likelihood = -0.32 | Transmembrane | 342 - 358 ( 342 - 360) |

----- Final Results -----  
           bacterial membrane --- Certainty=0.4609(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-2484-

A related GBS nucleic acid sequence <SEQ ID 9397> which encodes amino acid sequence <SEQ ID 9398> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB14651 GB:Z99117 amino acid permease [Bacillus subtilis]  
Identities = 165/361 (45%), Positives = 227/361 (62%), Gaps = 17/361 (4%)

Query: 1 MGIFLT-LSYWISLIFIGMAEITAVGEYVQFWFPEWPSWIIQIVFLAILSSINLIAVKAF 59  
M F+T +YW I + MA++TAVG Y Q+W P+ P W+ ++ L IL +NL VK F

10 Sbjct: 95 MAAFITGWTYWFCWISLAMADLTAVGIYTYWLPDVPQWLPGLLALIILLIMNLATVKLF 154

Query: 60 GETEFWFAMIKVIAAILGLIATGIFMVLTNFDTGHGYHASISNITNHFWEFPPKGLNFFMA 119  
GE EFWFA+IKVIAIL LI TGI ++ F G AS++N+ +H FP G F ++

15 Sbjct: 155 GELEFWFALIKVIAIALIVTGILLIAKGFSASG--PASLNNLWSHGGMFPNGWHGFILS 213

Query: 120 FQMVFAYLAIEFVGVTSETANPRKVLPAKIQEIPMRIILFYAGSLLAIMAIFPWQQLP 179  
FQMV FA++ IE VG+T ET NP+KV+PKAI +IP+RI+LFY G+L IM I+PW L

Sbjct: 214 FQMVFVFAFVGIELVGLTAGETENPQKVIPKAINQIPVRILLFYVGALFVIMCIYPWNVLN 273

20 Query: 180 VNESPFTVFKLAGIKWAAALINFFVLTSAAALNSTLYSTGRHLFQLANE--SPNALTK 237  
NESPFV VF GI AA+LINFVVLTSAAAS NS L+ST R ++ LA + +P L K

Sbjct: 274 PNESPFTVQVFSAVGIVVAASLINFVVLTSAAASALFSTSRMVSLAKDHHAPGLLKK 333

25 Query: 238 ALKLDQLSRQSVPSRAIIAS--AVIVGASALISVLPGISDAFSLITASSSGVYISYVLI 295  
L+ +VPS A+ S A+++G S L ++P F+LIT+ S+ +I I+ +

Sbjct: 334 -----LTSSNVPSNALFFSSIAILIGVS-LNYLMP--EQVFTLITSVSTICFIFIWGIT 384

Query: 296 MIAHWKYRKS--PDFMEDGYKMPAYKILSPITLLFFLFVFSVSLFLQDSTYIGAIGATIWI 354  
+I H KYRK+ + + +KMP Y + + +TL F F+ V L L + T I +W +

30 Sbjct: 385 VICHLYRKYRQHEAKANKFKMPFYPLSNYLTALFLAFLILVILALANDTRIALFVTPVWFV 445

There is also homology to SEQ ID 4070:

Identities = 286/364 (78%), Positives = 322/364 (87%), Gaps = 1/364 (0%)

35 Query: 2 GIFLTLSYWISLIFIGMAEITAVGEYVQFWFPEWPSWIIQIVFLAILSSINLIAVKAFGE 61  
G F LSYWISLIFIGMAEITAVG YVQFWFP WP+W+IQ+VFL +LSSINLIAV+ FGE

Sbjct: 101 GYFSGLSYWISLIFIGMAEITAVGAYVQFWFSPWPAWLIQLVFLVLLSSINLIAVRVFE 160

40 Query: 62 TEFWFAMIKVIAAILGLIATGIFMVLTNFDTGHGYHASISNITNHFWEFPPKGLNFFMAFQ 121  
TEFWFAMIK++AIL LIAT IFMVL T F+T H HAS+SNI +HF FP GKL FFMFAFQ

Sbjct: 161 TEFWFAMIKILAILALIALIATAIFMVLTFET-HTGHASLSNIFDHFSMFNGKLKFFMAFQ 219

Query: 122 MVFFAYLAIEFVGVTSETANPRKVLPAKIQEIPMRIILFYAGSLLAIMAIFPWQQLPVN 181  
MVFFAY AIEFVG+TTSETANPRKVLPAKIQEIP RI++FY G+L++IMAI PW QLPV+

45 Sbjct: 220 MVFFAYQAIEFVGITSETANPRKVLPAKIQEIPTRIVIFYVGALVSIMAIVPWHQLPVD 279

Query: 182 ESPFVTVFKLAGIKWAAALINFFVLTSAAALNSTLYSTGRHLFQLANESPNALTKALKL 241  
ESPFV VFKL GIKWAAALINFFVLTSAAALNSTLYSTGRHL+Q+ANE+PNALT LK+

50 Sbjct: 280 ESPFVMVFKLIGIKWAAALINFFVLTSAAALNSTLYSTGRHLYQIANETPNALTNRLKI 339

Query: 242 DQLSRQSVPSRAIIASAVIVGASALISVLPGISDAFSLITASSSGVYISYVLIIMIAHWK 301  
+ LSRQ VPSRAIIASAV+VG SALI++LPG++DAFSLITASSSGVYI+IY L MIAHWK

Sbjct: 340 NTLSRQGVPSRAIIASAVVVGISALINILPGVADAFSLITASSSGVYIAIYALTMIAHWK 399

55 Query: 302 YRKSPDFMEDGYKMPAYKILSPITLLFFLFVFSVSLFLQDSTYIGAIGATIWIIGFGLYSH 361  
YR+S DFM DGY MP YK+ +P+TL FF FVF+SLFLQ+STYIGAIGATIWI FG+YS+

Sbjct: 400 YRQSKDFMADGYLMPKYKVTTPLTALFAFVVFISLFLQESTYIGAIGATIWIIFGIYSN 459

Query: 362 FKHK 365  
K K

60 Sbjct: 460 VKFK 463

-2485-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2205

A DNA sequence (GBSx2324) was identified in *S.agalactiae* <SEQ ID 6807> which encodes the amino acid sequence <SEQ ID 6808>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have an uncleavable N-term signal seq

|          |                    |               |                        |
|----------|--------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -8.33 | Transmembrane | 194 - 210 ( 191 - 215) |
| INTEGRAL | Likelihood = -5.47 | Transmembrane | 17 - 33 ( 14 - 38)     |
| INTEGRAL | Likelihood = -5.15 | Transmembrane | 125 - 141 ( 119 - 144) |
| INTEGRAL | Likelihood = -3.88 | Transmembrane | 155 - 171 ( 153 - 176) |
| INTEGRAL | Likelihood = -1.38 | Transmembrane | 96 - 112 ( 94 - 114)   |
| INTEGRAL | Likelihood = -0.43 | Transmembrane | 49 - 65 ( 49 - 65)     |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.4333(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]

Identities = 80/214 (37%), Positives = 122/214 (56%), Gaps = 3/214 (1%)

Query: 4 FFSNIRTEIPQMPLLIHSLILSVLPFLMWLTLVNRDKPLYKTIWSILLGLQLITIYTWWF 63

FF+ T+ P+ L + + ++L + R+K +Y+ + IL +QLI +Y W++

Sbjct: 7 FFTTQATKPPKFDLFWYVSLFTLLALTFYTAHRYREKKVYQRFQILQTVQLLILLYGWYW 66

Query: 64 WAKLPLSESLPLYHCRIGMEFVLLARPGI--LKDYFALLGVVGGVLAAMIHPDFYPYQFLH 121

+PLSESLP YHCR+ MFVVLL PG K YFALLG G + A ++P Y F H

Sbjct: 67 VNHMPLSESLPFYHCRAMFVVLL--PGQSKYKQYFALLGTFGTAAAFVYVPDAYPPPH 125

Query: 122 VTNIFFFIHGFALFVLSLLHLMTQSNLDKLNPKLIITLLINMSLIFINLLTGGNYGFM 181

+T + F GH AL SL++L+ Q N L+ K I +T +N + +NL+TGG+YGF+

Sbjct: 126 ITILSFIFGHLALLGNSLVYLLRQYNARLLDVKGIFLMTFALNALIFVNVNLTGGDYGFL 185

Query: 182 MKTPILGITNPFLNLFIVTTLSSFLVLFVKQIFQ 215

K P++G N +V+ +L + K+I +

Sbjct: 186 TKPPLVGDHGLVANYLLVSIVLVATISLTKKILE 219

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6809> which encodes the amino acid sequence <SEQ ID 6810>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

|          |                     |               |                        |
|----------|---------------------|---------------|------------------------|
| INTEGRAL | Likelihood = -11.25 | Transmembrane | 16 - 32 ( 11 - 39)     |
| INTEGRAL | Likelihood = -3.45  | Transmembrane | 154 - 170 ( 153 - 173) |
| INTEGRAL | Likelihood = -3.08  | Transmembrane | 96 - 112 ( 94 - 112)   |
| INTEGRAL | Likelihood = -1.91  | Transmembrane | 191 - 207 ( 191 - 209) |
| INTEGRAL | Likelihood = -1.12  | Transmembrane | 71 - 87 ( 71 - 87)     |

----- Final Results -----

|                     |     |                               |         |
|---------------------|-----|-------------------------------|---------|
| bacterial membrane  | --- | Certainty=0.5501(Affirmative) | < succ> |
| bacterial outside   | --- | Certainty=0.0000(Not Clear)   | < succ> |
| bacterial cytoplasm | --- | Certainty=0.0000(Not Clear)   | < succ> |

The protein has homology with the following sequences in the databases:

>GP:AAC95438 GB:AF068901 unknown [Streptococcus pneumoniae]

Identities = 90/231 (38%), Positives = 128/231 (54%), Gaps = 7/231 (3%)

Query: 3 FFAIDPIGLPHTSLIFYLSSLLIALLLVFLTFQAYRLKS-HRYFFLFLQLSQVIGLYTWY 61

-2486-

FF P L +Y+S L L L F T YR K. ++ FF LQ Q+I LY WY  
 Sbjct: 7 FFTTQATKPPKFDLFWYVS-LFTLLALTFYTAHRYREKKVYQRFFQILQTVQLILLYGWY 65  
 Query: 62 VLRGFPLDEALPLYHCRAMLAIFFLPDRNKFQLEFMVLGIGGTFLALL--SPDLYPFRL 119  
 + PL E+LP YHCR+AM + LP ++K+KQ F +LG GT A + PD YPF  
 Sbjct: 66 WVNHMPLSESLPFYHCRMAMFVLLLPQSKYKQYFALLGTFTLAAAFVYPVPDAYPPF- 124  
 Query: 120 WHVANVSFYFGHYALLVNGLIYLLRFYDASQLRLLSVVRYLATVNFLLLLVSLATKGNYG 179  
 H+ +SF FGH ALL N L+YLLR Y+A L + + +N L+ +V+L T G+YG  
 Sbjct: 125 -HITILSFIFGHLALLGNSLVLLRQYNARLLDVKGIFLMTFALNALIFVNLVTGGDYG 183  
 Query: 180 FVMDIPVIHTRHLLLNFIIVTSGLTFMVKITEYFYLFKFGQAQQLALAFSKE 230  
 F+ P++ L+ N+++V+ L + +T+ L+F AQ+ KE  
 Sbjct: 184 FLTKPPLVGDHGLVANYLLVSIVLVATISLTKKI-LEFFLAQEAEMIVKE 233

An alignment of the GAS and GBS proteins is shown below.

Identities = 70/216 (32%), Positives = 117/216 (53%), Gaps = 1/216 (0%)

Query: 2 IEFFSNIRTEIPQMPLLIHSLILSVLPFLMWLTLVNRDKPLYKTIWSILLGLQLITYTW 61  
 ++FF+ +P L+ + L+ L++LT ++ + L Q+I +YTW  
 Sbjct: 1 MDFFAIDPIGLPHTSLIFYLSSLLIALLLVFLTFQAYRLKSHRYFFLFLQLSQVIGLYTW 60  
 Query: 62 FFWAKLPLSESLPLYHCRIGMFVVL-LARPGILKDYFALLGVVGGVLAMIHPDFYPYQFL 120  
 + PL E+LPLYHCRI M + L K F +LG+ G LA++ PD YP++  
 Sbjct: 61 YVLRGFPLDEALPLYHCRAMLAIFFLPDRNKFQLEFMVLGIGGTFLALLSPDLYPFRLW 120  
 Query: 121 HVTNIFFFIGHFALFVLSLLHLMTQSNLDKLNPKLIQLTLLINMSLIFINLLTGGNYGF 180  
 HV N+ F+ GH+AL V L++L+ + +L +++ +N L+ ++L T GNYGF  
 Sbjct: 121 HVANVSFYFGHYALLVNGLIYLLRFYDASQLRLLSVVRYLATVNFLLLLVSLATKGNYG 180  
 Query: 181 MMKTPILGITNPFLNLFIVTTLTSLFVLVFKQIFQK 216  
 +M P++ + LN IVT+ L+F+V + + K  
 Sbjct: 181 VMDIPVIHTRHLLLNFIIVTSGLTFMVKITEYFYLK 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2206

A DNA sequence (GBSx2325) was identified in *S.agalactiae* <SEQ ID 6811> which encodes the amino acid sequence <SEQ ID 6812>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3297(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2207

A DNA sequence (GBSx2326) was identified in *S.agalactiae* <SEQ ID 6813> which encodes the amino acid sequence <SEQ ID 6814>. This protein is predicted to be oxalate:formate antiporter (oxIT-2). Analysis of this protein sequence reveals the following:



-2487-

Possible site: 27

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

|    |          |                    |               |                        |
|----|----------|--------------------|---------------|------------------------|
|    | INTEGRAL | Likelihood = -7.80 | Transmembrane | 380 - 396 ( 376 - 399) |
|    | INTEGRAL | Likelihood = -7.43 | Transmembrane | 291 - 307 ( 284 - 310) |
| 5  | INTEGRAL | Likelihood = -5.63 | Transmembrane | 169 - 185 ( 163 - 186) |
|    | INTEGRAL | Likelihood = -4.99 | Transmembrane | 226 - 242 ( 223 - 245) |
|    | INTEGRAL | Likelihood = -4.19 | Transmembrane | 46 - 62 ( 39 - 63)     |
|    | INTEGRAL | Likelihood = -4.09 | Transmembrane | 311 - 327 ( 308 - 329) |
|    | INTEGRAL | Likelihood = -1.49 | Transmembrane | 261 - 277 ( 260 - 278) |
| 10 | INTEGRAL | Likelihood = -1.06 | Transmembrane | 133 - 149 ( 133 - 150) |
|    | INTEGRAL | Likelihood = -0.85 | Transmembrane | 98 - 114 ( 98 - 114)   |
|    | INTEGRAL | Likelihood = -0.06 | Transmembrane | 77 - 93 ( 77 - 93)     |

----- Final Results -----

|    |                                                              |
|----|--------------------------------------------------------------|
| 15 | bacterial membrane --- Certainty=0.4121(Affirmative) < succ> |
|    | bacterial outside --- Certainty=0.0000(Not Clear) < succ>    |
|    | bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>  |

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus  
lactis]  
Identities = 220/398 (55%), Positives = 306/398 (76%), Gaps = 3/398 (0%)

25 Query: 5 NRYVVAVSGVVLHMLGSTYAWSVFRNPIISETGWDISSVSFAFSLAIFCLGMSAAFMGH 64  
NRYVVA +GV+ HLM+GS YAWSVF NPI + GW SSV+ AFS+AI+ LGMSAAFMG  
Sbjct: 4 NRYVVAFAAGVMFHLMIGSVYAWSVFTNPIAKQNGWAESSVALAFSAIYFLGMSAAFMGK 63

Query: 65 LVERFGPRIMGMISAILYGAGNVLTGLAIETQQLWLLYVAYGILGGIGLGSYITPVSTI 124  
+VE+ GPR+ G I++ LYG G ++TG AI +WLLY++YG++GG+GLG+GY+TPVSTI  
30 Sbjct: 64 VVEKIGPRLTGTIASFLYGTGTIMTGWAIHQNSIWLlyLSYGVIGGLGLGAGYVTPVSTI 123

Query: 125 IKWFPDRRLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMIASQF 184  
IKWFPD+RGLATG AIMGFGFA+++T P+AQ LM +G+ +TFY+LG YF +M++A+QF  
35 Sbjct: 124 IKWFPDKRGLATGLAIMGFGFAAMLTPVAQQLMASVGLGQTFYLLGTFYFVIMLLAAQF 183

Query: 185 IKQPPQEKITILTHDQKKNAMNSQIITG--LKANAIAKSKTFYIIWLTFLFINISCGGLGI 242  
I + P ++ T + +++ G L AN A+K+K+F +W+ FINI+CG+GL+  
Sbjct: 184 IVR-PNLALSSTTENSISQKKGTRLTRGPBLTANQALKTKSFTFLWIMFFINITCGIGLV 242

40 Query: 243 SAASPMAQDLAGYSAESAALLVGVLGIFNGFGRLWASLSDYIGRPLTFFILFIVNFIMT 302  
SAASPMAQ + G S ++AA++VG++G+FNFGRL+WA+LSDYIGRP TF +FI++ +M  
Sbjct: 243 SAASPMAQSMTGMSVQTAAIMVGIIGLFNGFGRLIATLSDYIGRPATFSAIFILDIVML 302

45 Query: 303 SSLFLSFNAIVFAIAMSILMTCYAGAGFSLLPAYLSDIFGKELATLHGYSLTAWAIAGLF 362  
S++ + ++F IA+ +LM-CYAGAGF++PAYL D+FGTKEL +HGY LTAWA AG+  
Sbjct: 303 SAILIFKPLPLLFVIALCLLMSCYAGAGFSVIPAYLGDVFGTKELGAVHGYVLTAWAAAGV 362

Query: 363 GPLLLSKTYSWGSNSYQLTLMVFGFLFGLLLSLYLRLK 400  
GPLLLS T+ ++Y LTL F + L LL+S ++++  
50 Sbjct: 363 GPLLLSLTHQLFHNYYTLTAAFILIDLLALLISFWIQR 400

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6815> which encodes the amino acid sequence <SEQ ID 6816>. Analysis of this protein sequence reveals the following:

Possible site: 27

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

|    |          |                     |               |                        |
|----|----------|---------------------|---------------|------------------------|
|    | INTEGRAL | Likelihood = -12.95 | Transmembrane | 289 - 305 ( 282 - 321) |
|    | INTEGRAL | Likelihood = -11.83 | Transmembrane | 376 - 392 ( 372 - 397) |
|    | INTEGRAL | Likelihood = -8.55  | Transmembrane | 163 - 179 ( 160 - 189) |
|    | INTEGRAL | Likelihood = -7.75  | Transmembrane | 227 - 243 ( 221 - 247) |
| 60 | INTEGRAL | Likelihood = -5.89  | Transmembrane | 44 - 60 ( 41 - 67)     |
|    | INTEGRAL | Likelihood = -1.38  | Transmembrane | 310 - 326 ( 309 - 327) |
|    | INTEGRAL | Likelihood = -0.90  | Transmembrane | 353 - 369 ( 353 - 369) |
|    | INTEGRAL | Likelihood = -0.37  | Transmembrane | 138 - 154 ( 138 - 154) |
|    | INTEGRAL | Likelihood = -0.06  | Transmembrane | 98 - 114 ( 98 - 114)   |
| 65 | INTEGRAL | Likelihood = -0.00  | Transmembrane | 259 - 275 ( 259 - 275) |

-2488-

## ----- Final Results -----

5 bacterial membrane --- Certainty=0.6180(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF36228 GB:AF168363 oxalate:formate antiporter [Lactococcus  
 lactis]  
 10 Identities = 222/399 (55%), Positives = 305/399 (75%), Gaps = 3/399 (0%)

Query: 3 KTKRYIIATAGILLHMLGSTYAWSVYRNPIQETGWDQAPVAFSLAIFCLGLSAAFM 62  
 KT RY++A AG++ HLM+GS YAWSV+ NPI ++ GW ++ VA AFS+AI+ LG+SAAFM  
 15 Sbjct: 2 KTNRYVAVAFAGVMFHLMIGSVYAWSVFTNPIAKQNGWAESSVALAFSIAIYFLGMSAAFM 61

Query: 63 GNLVEQYGPRLTGTVSAILYASGNMLTGLAIDRKEIWLILYIGYGVIGGLGLGAGYITPIS 122  
 G +VE+ GPRLTGT+++ LY +G ++TG AI + IWLLY+ YGVIGGLGLGAGY+TP+S  
 Sbjct: 62 GKVVKEIGPRLTGTITASFLYGTGTIMTGWAIHQNSIWLLYLSYGVIGGLGLGAGYVTPVS 121

20 Query: 123 TTIKWFPDKRGMATGFAIMGFGFASLLTSPIAQWLIETEGLVATFYLLGLIYLIVMLFAS 182  
 TTIKWFPDKRG+ATG AIMGFGFA++LT P+AQ L+ + GL TFYLLG Y ++ML A+  
 Sbjct: 122 TTIKWFPDKRGLATGLAIMGFGFAAMLTPVAQQLMASVGLEQTFYLLGTFYFVIMLLAA 181

25 Query: 183 QLIHKPTAAEIAILDKKRLQ--NNSYLIEG--MTAKEALKTKSFYCLWVILFINITCGLGL 239  
 Q I++P A + + Q + L G +TA +ALKTKSF LW++ FINITCG+GL  
 Sbjct: 182 QFIVRPNALSSSTTENSISQKKGTRLTGRPELTANQALKTKSFTFLWIMFFINITCGIGL 241

Query: 240 ISVVAPMAQDLTGMSPEMSAIVVGAMGIFNGFGRLVWASLSDYIGRRVTVILLFLVSIIM 299  
 +S +PMAQ +TGMS + +AI+VG +G+FNGFGRL+WA+LSDYIGR T +F++ I+M  
 30 Sbjct: 242 VSAASPMASMTGMSVQTAAIMVGIIGLFNGFGRLIWATLSDYIGRPATFSAIFILDIVM 301

Query: 300 TISLIFAHSSILFIMISIATLMTCYGAGFSLIPPYLSDLFGAKELATLHGYYLTAWAIAAL 359  
 +++ L+F+I++ LM+CYGAGFS+IP YL D+FG KEL +HGY+LTAWA A +  
 35 Sbjct: 302 LSAILIFKLPLLFVIALCLLMSCYGAGFSVIPAYLGDVFGTKELGAVHGYVLTAWAAAGV 361

Query: 360 TGPMLLSITVEWTHNYLLTLCVFIVLYILGLMVALRLKK 398  
 GP+LLS+T + HNY LTL FI++ +L L+++ +++  
 Sbjct: 362 VGPLLLSLTHQLFHNYYTLTAAFILIDLALLISFWIQR 400

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 252/400 (63%), Positives = 329/400 (82%), Gaps = 2/400 (0%)

Query: 1 MKNLNRYVAVSGVVLHMLGSTYAWSVFRNPIISETGWDISSVSFAFSLAIFCLGMSAA 60  
 M+ RY++A +G++LHMLGSTYAWSV+RNPI+ ETGWD + V+FAFSLAIFCLG+SAA  
 45 Sbjct: 1 MEKTKRYIIATAGILLHMLGSTYAWSVYRNPIQETGWDQAPVAFSLAIFCLGLSAA 60

Query: 61 FMGHLVERFGPRIMGISAILYAGNVLGTGLAIETQQLWLLYVAYGILGGIGLGSYITP 120  
 FMG+LVE++GPR+ G +SAILY +GN+LTGLAI+ +++WLLY+ YG++GG+GLG+GYITP  
 50 Sbjct: 61 FMGNLVEQYGPRLTGTVSAILYASGNMLTGLAIDRKEIWLILYIGYGVIGGLGLGAGYITP 120

Query: 121 VSTIIKWFPDRRLATGFAIMGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMMI 180  
 +STIIKWFPD+RG+ATGFAIMGFGFASL+TSP+AQ L+ G+ TFY+LGL+Y VM+  
 Sbjct: 121 ISTIIKWFPDKRGMATGFAIMGFGFASLLTSPIAQWLIETEGLVATFYLLGLIYLIVMLF 180

55 Query: 181 ASQFIKQPQEKITILTHDGKKNAMNSQIITGLKANAAIKSKTFYIIWLTFLFINISCGLG 240  
 ASQ I +P +I IL D K+ NS +I G+ A A+K+K+FY +W+ LFINI+CGLG  
 Sbjct: 181 ASQLIIKPTAAEIAIL--DKKRLQNNNSYLIEGMTAKEALKTKSFYCLWVILFINITCGLG 238

Query: 241 LISAASPMADLAGYSAESAALVGVLGIFNGFGRLWASLSDYIGRPLTFIILFIVNFI 300  
 LIS +PMAQDL G S E +A++VG +GIFNGFGRL+WASLSDYIGR +T I+LF+V+ I  
 60 Sbjct: 239 LISVVAPMAQDLTGMSPEMSAIVVGAMGIFNGFGRLVWASLSDYIGRRVTVILLFLVSII 298

Query: 301 MTSSLFLSFNAIVFAIAMSILMTCYGAGFSLPAYLSDFGTGKELATLHGYSLTAWAIAAG 360  
 MT SL + ++++F I+++ LMTCYGAGFSL+P YLSD+FG KELATLHGYS LTAWAIA  
 65 Sbjct: 299 MTISLIFAHSSILFIMISIATLMTCYGAGFSLIPPYLSDLFGAKELATLHGYYLTAWAIAA 358

Lipop: Possible site: -1   Crend: 5  
McG: Discrim Score:       5.06  
GvH: Signal Score (-7.5): 4.38

```

Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program      count: 10 value: -7.80 threshold: 0.0
INTEGRAL Likelihood = -7.80 Transmembrane 380 - 396 ( 376 - 399)
INTEGRAL Likelihood = -7.43 Transmembrane 291 - 307 ( 284 - 310)
INTEGRAL Likelihood = -5.63 Transmembrane 169 - 185 ( 163 - 186)
INTEGRAL Likelihood = -4.99 Transmembrane 226 - 242 ( 223 - 245)
INTEGRAL Likelihood = -4.19 Transmembrane 46 - 62 ( 39 - 63)
INTEGRAL Likelihood = -4.09 Transmembrane 311 - 327 ( 308 - 329)
INTEGRAL Likelihood = -1.49 Transmembrane 261 - 277 ( 260 - 278)
INTEGRAL Likelihood = -1.06 Transmembrane 133 - 149 ( 133 - 150)
INTEGRAL Likelihood = -0.85 Transmembrane 98 - 114 ( 98 - 114)
INTEGRAL Likelihood = -0.06 Transmembrane 77 - 93 ( 77 - 93)
PERIPHERAL Likelihood = 0.42 352
modified ALOM score: 2.06

```

\*\*\* Reasoning Step: 3

```

----- Final Results -----
      bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

35 ORF02272(313 - 1500 of 1818)  
GP|7107009|gb|AAF36228.1|AF168363\_4|AF168363(4 - 400 of 421) oxalate:formate antiporter  
{*Lactococcus lactis*}  
%Match = 38.5  
%Identity = 55.4 %Similarity = 79.1  
Matches = 220 Mismatches = 81 Conservative Sub.s = 94

GK\*IC\*AENW\*YIQFFDNLFITNYIFKNKT\*VRF\*EDCLKNLRNVVA VSGVLHMLGSTYAWSVFRNP I ISETGWDIS  
||||| :||::||:| ||||| || | : ||  
MKTNRVVA FAGVMFHL MIGSVYAWSVF TNP IAKQNGWAES  
10 20 30 40

SVSF AFSIAI FCLGMSAA FMGH LVER FGPR IMGMISA ILYGAG NVLTGLA IETQQLWLLY VAYGI LGGTGL GSGYITPV S  
||::|||::|: ||||||| :||: |||: | :: ||| | ::|| || :|||::|::|::|::|::|::|::|::|  
SVALAFSIAI YFLGMSAA FMKG VKVEKIG PRLTG TIASF LYGTGTIMTGWAI HQNSIWLLYLSYGVI GGGLGL GAGYVTPVS

60 70 80 90 100 110 120

696           726           756           786           816           846           876           906  
TIIKWFPDRRLGATGFAIMGGFGFASLVTSPLAQSLMIRIGVGKTFYILGLVYFFVMIMASQFIKQPPOKEKITILT HDGKK  
||||| |: ||||| : ||||| :: | : | | : | : ||| : | | : :: | : | | : | : : : | :  
TIIKWFDPKRGLATGLAIMGFGFAAML TGPVAQQLMASVGLEQT FYLLGT F YFVIMLLAAQFIVRP-NLALSSTENSIS  
               140           150           160           170           180           190           200

936            960            990            1020            1050            1080            1110            1140  
NAMNSQIITG--LKANAAIKSTFYIIWLTFLFINISCGGLGISAASPMADLAGYSAESAALLVGVLGIFNGFGRLIWAS  
:: | | | |:|:| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:  
QKKGTRLTRGPELTANQALKTKSFTEFLWMFFINITCGIGLVSAASPMASMTGMSVQTAAIMVGIIGLFNGFGRLIWA

210            220            230            240            250            260            270            280

1170      1200      1230      1260      1290      1320      1350      1380

-2490-

```

5      LSDYIGRPLTFIILFIVNFIMTSSLFSLFNAIVFAIAMSILMTCYAGFSLLPAYLSDIFGTEKELATLHGYSLTAWAIAG
      ||||| || :||: :| |::: :| ||: :||:|||||::||| |:||||| :||| ||||| ||
      LSDYIGRPATFSAIFILDIVMLSAILIFKLPFLFVIALCLMSCYAGFSPVLPAYLGDVFGTEKELGAVHGYYVLTAWAAAG
      290      300      310      320      330      340      350      360

10     1410      1440      1470      1500      1530      1560      1590      1620
      LFGPILLSKTYSWGNSYQLTLMVFGFLFLFGLLSLYLRKLTTKV*YISNLKFFGFTKEFFL*KIVLSYSK*FDILSI*
      : ||||| |: :| ||| | :| ||:|:::
      VVGPLLSLTHQLFHNYYTLAAFILIDLALLISFWIQRDFIKASKLIKKQIKNYFKAH
      370      380      390      400      410      420

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2208

15 A DNA sequence (GBSx2327) was identified in *Sagalactiae* <SEQ ID 6817> which encodes the amino acid sequence <SEQ ID 6818>. This protein is predicted to be D-Ala-D-Ala adding enzyme (murF). Analysis of this protein sequence reveals the following:

```

20     Possible site: 45
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1311(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9739> which encodes amino acid sequence <SEQ ID 9740> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

30     >GP:AAC95436 GB:AF068901 D-Ala-D-Ala adding enzyme [Streptococcus pneumoniae]
      Identities = 313/453 (69%), Positives = 375/453 (82%)

      Query: 32  MKLSLHEVAKVVGAKNQVSEFEDVPLGNIEFDSRNISEGDLFLPLKGARDGHEFIEMAFD 91
      MKL++HE+A+VVGAKN +S FED L EFDSR I GDLF+PLKGARDGH+FIE AF+
      Sbjct: 1   MKLTIHEIAQVVGAKNDISIFEDTQLEKAEFDSRLIGTGDLFVPLKGARDGHDFIETAFE 60

35     Query: 92  NGAIATISEKEIEGHPYLLVSDALKAFQVLAQYYIEKMNVDVIAVTGSNGKTTTKDMIAA 151
      NGA T+SEKE+ HPY+LV D L AFQ LA YY+EK VDV AVTGSNGKTTTKDM+A
      Sbjct: 61  NGAAVTLSEKEVSNHPYIILVDDVLTAFAQSLASYLEKTTVDVFAVTGSNGKTTTKDMLAH 120

40     Query: 152 ILSTYKTYKTQGNYNNEIGLPYTVLHMPEDTEKIILEMGQDHLGDIHVLSEIAKPRIAV 211
      +LST YKTYKTQGNYNNEIGLPYTVLHMPE TEK++LEMGQDHLGDIH+LSE+A+P+ A+
      Sbjct: 121 LLSTRYKTYKTQGNYNNEIGLPYTVLHMPEGTEKLVLEMGQDHLGDIHLLSELARPKTAI 180

45     Query: 212 VTLIGEAEHLEFFGSREKIAEGKMQITDGMSSDGILIA PGDPIIDPYLPANQMTIRFGHDQ 271
      VTL+GEAHL FF R +IA+GKMQI DGM+S +L+AP DPI++ YLP ++ +RFG
      Sbjct: 181 VTLVGEAHLAFFKDRSEIAKGKMQIADGMASGSLLLAPADPIVEDYLPDCKKVVRFQGQA 240

50     Query: 272 ELQVTELKEEKHSLTFKTNALHQLRIPVPGKYNATNAMVAAYVGKLLAVAEEEDIVDALE 331
      EL++T+L E K SLTFK N LE L +PV GKYNATNAM+A+YV V+EE I A +
      Sbjct: 241 ELEITDLVERKDSLTFKANFLEQVLDLPVTGKYNATNAMIASYVALQEGVSEEQIHQAFQ 300

55     Query: 332 NLQLTRNRTEWKKSAANGADILSDVYNANPTAMRLILETFSAIPNNDGGKKIALLADMKEL 391
      +L+LTRNRTEWKK+ANGADILSDVYNANPTAM+LILETFSAIP N+GGKKIA+LADMKEL
      Sbjct: 301 DLELTRNRTEWKKAAANGADILSDVYNANPTAMKLILETFSAIPANEGGKKIAVLADMKEL 360

      Query: 392 GEQSVDLHNQMIMSIRPDSIDTLICYQDIEGLAQLASQMFPIGKVYFFKKNQEVDDQFDQ 451
      G QSV LHNQMI+S+ PD +DT+I YG+DI LAQLASQMFPIG VY+FKK ++ DQF+
      Sbjct: 361 GNQSVQLHNQMILSLSPDVLDTVIFYGEDIAELAQLASQMFPIGHVYFFKKTEDQDQFED 420

```

-2491-

Query: 452 LLAKVKDTLKEKDQILLKGSNSMNLISKIVDILE 484  
 L+ +VK++L DQILLKGSNSMNL+ +V+ LE  
 Sbjct: 421 LVKQVKESLSANDQILLKGSNSMNLAMLVESLE 453

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6819> which encodes the amino acid sequence <SEQ ID 6820>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 323/452 (71%), Positives = 387/452 (85%)

Query: 32 MKLSLHEVAKVVGAKNQVSEFEDVPLGNIEFDSRNISEGDLFLPLKGARDGHEFIEMAFD 91  
 MKL+LHEVAK+V A+N VS+ +DVPL +IEFDSR I++GDLEFLPLKG RDGHEFI++AF  
 20 Sbjct: 1 MKLTLHEVAKIVDAQNNVSDLDVPLHHIEFDSRKITKGDLFLPLKGQRDGHFIDLAQF 60

Query: 92 NGAIATISEKEIEGHPYLLVSDALKAFQVLAQYYIEKMNVDIVIAVTGSNGKTTTKDMIAA 151  
 NGA+AT SEKE+ G P+LLV D LKAFQ LA YYI+KM VDVIATGSNGKT+TKDMI A  
 25 Sbjct: 61 NGAVATFSEKELPGKPHLLVEDCLKAFQKLAHYIIDKMRVDVIAVTGSNGKTSTKDMIGA 120

Query: 152 ILSTTYKTYKTQGNYNNEIGLPYTVLHMPEDTEKIILEMGQDHLGDIHVLSEIAKPRIAV 211  
 +LSTTYKTYKTQGNYNNEIGLPYTVLHMP+DTEKI+LEMGQDH+GDI +LSEIA+PRIAV  
 Sbjct: 121 VLSTTYKTYKTQGNYNNEIGLPYTVLHMPDDTEKIVLEMGQDHMGDIRLLSEIARPRIAV 180

Query: 212 VTLIGEAHLEFFGSREKIAEGKMQITDGMSSDGILIAPGDPIIDPYLPANQMTIRFGHDQ 271  
 +TL+GEAHLE+FGSR+KIA+GKMQI DGM+SDGILIAPGDPIIDPYLP NQM IREG+ Q  
 30 Sbjct: 181 LTLVGEAHLEYFGSRDKIAQGKMQIVDGMNSDGILIAPGDPIIDPYLPENQMVIRFGNQ 240

Query: 272 ELQVTELKEEKHSLTFKTNALHQLRIPVPGKYNATNAMVAAYVGKLLAVAEEDIVDALE 331  
 E+ VT ++E+K SLTF TN L + +P+PGKYNATNAMVAAYVGKLLAV +EDI+ AL+  
 35 Sbjct: 241 EIDVTGIQEDKDSLTFITNVLATPVSILPLPGKYNATNAMVAAYVGKLLAVTDEDIIALQ 300

Query: 332 NLQLTRNRETEWKSANGADILSDVYNANPTAMRLILETFSAIPNNDGGKKIALLADMKEL 391  
 + LT NRTEWKK+ANGADILSDVYNANPTAMRLILETF+ I N GGKKIA+LADMKEL  
 40 Sbjct: 301 TVTLTGNRTEWKKANGADILSDVYNANPTAMRLILETFANIAKNPGKKIAVLADMKEL 360

Query: 392 GEQSVDLHNQMIMSIRPDSIDTLICYGQDIEGLAQLASQMFFIGKVYFFKKNQEVDDQFDQ 451  
 G+ SV LH+Q+I S+ +ID L+ YG I+ LA+LASQ++P +V++F K ++ DQF+  
 45 Sbjct: 361 GKDSVILHSQILDSLTSIGNIDQLVFYGDHIKELARLASQVYPAEQVHYFLKTEQEDQFEA 420

Query: 452 LLAKVKDTLKEKDQILLKGSNSMNLISKIVDIL 483  
 + V++ L DQILLKGS+SM+L K+VD L  
 Sbjct: 421 MAQYVQNILNPFDDQILLKGSMSLEKLVDR 452

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2209

A DNA sequence (GBSx2328) was identified in *S.agalactiae* <SEQ ID 6821> which encodes the amino acid sequence <SEQ ID 6822>. Analysis of this protein sequence reveals the following:

55 Possible site: 17  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1381(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-2492-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95435 GB:AF068901 D-Ala-D-Ala ligase [Streptococcus pneumoniae]  
Identities = 243/346 (70%), Positives = 289/346 (83%)

Query: 3 KETLILLYGGRSAEREVSLSAESVMRAINNDKFFVKTYFITQVGQFIKTQEFDEMPSSD 62  
K+T+ILLYGGRSAEREVSLSAESVMRA+NYD+F VKT+FI+Q G FIKTQEF P +  
Sbjct: 2 KQTIILLYGGRSAEREVSLSAESVMRAVNYDRFTVKTFFISQSGDFIKTQEFSHAPGQE 61

Query: 63 EKLMNTNQTVLDKMRPSDIYDDNAIVFPVLHGPMGEDGSIQGFLEVL RMPYVGTNLS 122  
++LMTN+T+D DK V PS IY++ A+VFPVLHGPMGEDGSIQGFLEVL+MPYVG NILSS  
Sbjct: 62 DRLMTNETIDWDKKVAPSAIYEEGAVVFPVLHGPMGEDGSIQGFLEVL RMPYVGCNLS 121

Query: 123 SVAMDKITTKQVLATVGPVQVAYQTYFEGDDLEHAIKLSLETLSFPFVKPANMGSSVGI 182  
S+AMDKITTK+VL + G+ QV Y EGDD+ I E L++P+F KP+NMGSSVGI  
Sbjct: 122 SLAMDKITTKRVLESAGIAQVPYVAIVEGDDVTAKIAEVEEKLAYPVFTKPSNMGSSVGI 181

Query: 183 SKATDESSLRSALDLALKYDSRLIEQGV TAREIEVGILGNNDVKTTFFPGEVVKDVFYD 242  
SK+ ++ LR A+ LA +YDSR+L+EQGV AREIEVG+LGN DVK+T PGEVVKDV FYD  
Sbjct: 182 SKSENQEELRQALKLAFRYDSRVLEQGVNAREIEVGILLGNVDKSTLPGEVVKDVAFYD 241

Query: 243 YDAKYIDNKITMDIPAKVDEATMEAMRQYASKAFKAIGACGLSRCDFFLT KDQGFIFLNEL 302  
YDAKYIDNKITMDIPAK+ + + MRQ A AF+AIG GLSRCDF T G+IFLNEL  
Sbjct: 242 YDAKYIDNKITMDIPAKISDDVAVMRQNAETAFAIGGLGLSRCDFFYTDKGEIFLNEL 301

Query: 303 NTMPGFTQWSMYPLLWENMGLTYSDLIEKLVMLAKEMFEKRESHLI 348  
NTMPGFTQWSMYPLLW+NMG++Y +LIE+LV LAKE F+KRE+HLI  
Sbjct: 302 NTMPGFTQWSMYPLLWENMGLTYSDLIEKLVMLAKESFDKREAHLI 347

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4559> which encodes the amino acid sequence <SEQ ID 4560>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1451 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 261/348 (75%), Positives = 306/348 (87%)

Query: 1 MSKETLILLYGGRSAEREVSLSAESVMRAINNDKFFVKTYFITQVGQFIKTQEFDEMPSS 60  
MSK+TL+LLYGGRSAEREVSLSAESVMRA+NYDKF VKTYFITQ+GQFIKTQ+F E PS  
Sbjct: 1 MSKQTLVLLYGGRSAEREVSLSAESVMRAVNYDKFLVKTYFITQMGQFIKTQQFSEKPS 60

Query: 61 SDEKLMNTNQTVLDKMRPSDIYDDNAIVFPVLHGPMGEDGSIQGFLEVL RMPYVGTNLS 120  
E+LMTN+T++L + ++PSDIY++ A+VFPVLHGPMGEDGSIQGFLEVL RMPY+GTN++  
Sbjct: 61 ESERLMTNETIELTQKIKPSDIYEEGAVVFPVLHGPMGEDGSIQGFLEVL RMPYIGTNVM 120

Query: 121 SSSVAMDKITTKQVLATVGPVQVAYQTYFEGDDLEHAIKLSLETLSFPFVKPANMGSSV 180  
SSS+AMDKITTK+VL ++G+PQVAY Y +G DLE + +L L+FPFVKPANMGSSV  
Sbjct: 121 SSSIAMDKITTKRVLESIGIPQVAYTVYIDGQDLEACL VETLRLTFFP FVKPANMGSSV 180

Query: 181 GISKATDESSLRSALDLALKYDSRLIEQGV TAREIEVGILGNNDVKTTFFPGEVVKDVF 240  
GISKA + LR AI LAL YDSR+LIEQGV AREIEVG+LGN+ VK+T PGEV+KDVF  
Sbjct: 181 GISKAQTKVELRKAIQLALTYDSRVLEIQGVVAREIEVGILLGNKVKSTLPGEVVKDVF 240

Query: 241 YDYDAKYIDNKITMDIPAKVDEATMEAMRQYASKAFKAIGACGLSRCDFFLT KDQGFIFLN 300  
YDY AKY+DNKITM IPA VD++ + MR YA AFKA+G CGLSRCDFFLT+DGQ++LN  
Sbjct: 241 YDYQAKYVDNKITMAIPADVDQSIIVTEMRSYAEVAFKALGGCGLSRCDFFLTQDGQVYLN 300

Query: 301 ELNTMPGFTQWSMYPLLWENMGLTYSDLIEKLVMLAKEMFEKRESHLI 348

-2493-

ELNTMPGFTQWSMYPLLWENMGL Y DLIE+LV LA+EMF++RESHLI  
 Sbjct: 301 ELNTMPGFTQWSMYPLLWENMGLAYPDLEEELVTLAQEMFDQRESHLI 348

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 5 vaccines or diagnostics.

### Example 2210

A DNA sequence (GBSx2329) was identified in *S.agalactiae* <SEQ ID 6823> which encodes the amino acid sequence <SEQ ID 6824>. This protein is predicted to be recombination protein (recR). Analysis of this protein sequence reveals the following:

10 Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 15 bacterial cytoplasm --- Certainty=0.2540(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:AAC44615 GB:U58210 RecM [Streptococcus thermophilus]  
 Identities = 181/198 (91%), Positives = 189/198 (95%)

Query: 1 MLYPTPIAKLIDSFSKLPGIGTKTATRLAFYTTIGMSDEDVNEFAKNLLAAKRELYCSVC 60  
 MLYPTPIAKLIDSFSKLPGIG KTATRLAFYTI MSDEDVN+FAKNLLAAKRELYCSVC  
 Sbjct: 1 MLYPTPIAKLIDSFSKLPGIGAKTATRLAFYTTISMSDEDVNDFAKNLLAAKRELYCSVC 60

25 Query: 61 GNLTDDEPCLICTDKTRDQSVILVVEDSKDVSAMEKIQEYNGLYHVLHGLISPMNGISPD 120  
 G LTDDDPCLICTD+TRD++ ILVVEDSKDVSAMEKIQEY GLYHVL GLISPMNG+ PD  
 Sbjct: 61 GRLTDDEPCLICTDETDRDKILVVEDSKDVSAMEKIQEYRGLYHVLQGLISPMNGVGPD 120

30 Query: 121 DINLKSILITRLMDGQVTEVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180  
 DINLKSILITRLMD +V EVI+ATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI  
 Sbjct: 121 DINLKSILITRLMDSEVDEVIIATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180

35 Query: 181 EYADEVTLLRAIENRTEL 198  
 EYADEVTLLRAIENRTEL  
 Sbjct: 181 EYADEVTLLRAIENRTEL 198

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6825> which encodes the amino acid sequence <SEQ ID 6826>. Analysis of this protein sequence reveals the following:

40 Possible site: 36  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 45 bacterial cytoplasm --- Certainty=0.2652(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 180/198 (90%), Positives = 192/198 (96%)

50 Query: 1 MLYPTPIAKLIDSFSKLPGIGTKTATRLAFYTTIGMSDEDVNEFAKNLLAAKRELYCSVC 60  
 +LYPTPIAKLIDS+SKLPGIG KTATRLAFYTTIGMS+EDVN+FAKNLLAAKRELYCS+C  
 Sbjct: 1 VLYPTPIAKLIDSYSKLPGIGIKTATRLAFYTTIGMSNEDVNDFAKNLLAAKRELYCSIC 60

55 Query: 61 GNLTDDEPCLICTDKTRDQSVILVVEDSKDVSAMEKIQEYNGLYHVLHGLISPMNGISPD 120  
 GNLTDDEPCLICTD +RDQ+ ILVVED+KDVSAAMEKIQEY+G YHVLHGLISPMNG+ PD  
 Sbjct: 61 GNLTDDEPCHICTDTSRDQTTILVVEDAKDVSAAMEKIQEYHGYHVLHGLISPMNGVGPD 120

-2494-

Query: 121 DINLKSLITRLMDGQVTEVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180  
 DINLKSLITRLMDG+V+EVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI  
 Sbjct: 121 DINLKSLITRLMDGKVSEVIVATNATADGEATSMYISRVLPAGIKVTRLARGLAVGSDI 180

5 Query: 181 EYADEVTLLRAIENRTEL 198  
 EYADEVTLLRAIENRTEL  
 Sbjct: 181 EYADEVTLLRAIENRTEL 198

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 10 vaccines or diagnostics.

### Example 2211

A DNA sequence (GBSx2330) was identified in *S.agalactiae* <SEQ ID 6827> which encodes the amino acid sequence <SEQ ID 6828>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3144(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 25 vaccines or diagnostics.

### Example 2212

A DNA sequence (GBSx2331) was identified in *S.agalactiae* <SEQ ID 6829> which encodes the amino acid sequence <SEQ ID 6830>. This protein is predicted to be penicillin-binding protein 2b. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood =-13.69 Transmembrane 23 - 39 ( 17 - 46)

----- Final Results -----  
 35 bacterial membrane --- Certainty=0.6477(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAC44614 GB:U58210 penicillin-binding protein 2b [Streptococcus thermophilus]  
 Identities = 341/683 (49%); Positives = 477/683 (68%), Gaps = 12/683 (1%)

Query: 4 RKKRYRLTVKKQNASIPRRNLNLLFFIIVLLFTVLILRLLEQMIGQQSFYMKKLTALTSYT 63  
 ++K R ++ +I RR+ LLF ++ +LF +L RL MQ+ +SFY KKL + YT  
 45 Sbjct: 18 KRKEKRANKPRKPVNISRRVYLLFGVVFLFLLLFARLTVMQVYNKSFYTKKLEDNSKYT 77

Query: 64 VKESKARGQIFDAKGVVLVENDERPTVAFSRGNNISSQSIKELANKLSHYITLTVASSD 123  
 V+ + RGQIFDAKG+ L N + + F+R N +SS ++K +A +L+ +TLTE +D  
 Sbjct: 78 VRIASERGQIFDAKGIALTTNQSKDVITFTRSNLVSSDTMKSVARLATLVTLTETKVID 137

50 Query: 124 RAKRDYYLADKANYKKVVESLPDSKRYDKFGNHLAESTVYANAVAAPVSAINYSEDELK 183  
 R KR++YLAD ANYK+VV LP+ K+ DKFGN LAE+T+Y NA+ AVP A++YSEDELK  
 Sbjct: 138 RQKREFYLADSANYKRVVNDLPNDKKTDFGNKLAETIYNNAINAVPDEAVDYSEDELK 197



-2495-

Query: 184 VVALFNQMNA TPTFGSVKLSTGELSDQIKKLDADKKELLGISVTSNWHRRKKGTSLSDI 243  
 +V +++ MNA F +V L T +L+ DQI + A +KEL GI V +W R +SLS +  
 Sbjct: 198 IVYIYSHMNAVSFSTVILKTADLTDPQIAIVA AKQKELNGIRVAKDWERHTSDSSLSPL 257

5 Query: 244 LGTISTEKAGLPREEVKYKKGYSLNDRVGTSYLEKQYEDDLQGIQIRKVVVNKKGKV 303  
 +G +S+ +AGLP+E+ K YLKKGY+LNDRVGTSYLEK+YE++LQG +R++ V+K+GKV  
 Sbjct: 258 IGRVSSSEAGLPQEDAKDYLLKKGYALNDRVGTSYLEKEYEEELQGKHTVREITVDKKGKV 317

10 Query: 304 VSDNITQEGKSGRNKLKLTIDLNYQNKVESILKQYYSSELSSGRASFSEGM YAVAIEPSTG 363  
 SD I Q+G G NLKLTIDL++Q VE IL Q SE+S +A++SEGM YAV + TG  
 Sbjct: 318 DSDKIITQKSGKGNLKLTLIDLPQKGVEDILGQQLSSEISGNKATYSEGM YAVVMNADTG 377

15 Query: 364 KVLAMAGLKNHDG--NLVDDSLGTIAKNFTPGSVVKGATLSSGWENKVLRGNEVLYDQEI 421  
 VLAMAG K++ G + D+LGTI FTPGSVVKGATL++GW + + G++VL DQ I  
 Sbjct: 378 AVLAMAGQKHEQGAQDFKADALGTITDVFTPGSVVKGATLTAGWRSGAIYGDQVLTDPQI 437

20 Query: 422 -----ANIRSWFT-RGLTPISAAQALEYSSNTYMQVALRLMGQDYNTGDALTDRGYQEA 475  
 I SWFT +G I+A QALEYSSNTYMQ+A++ +GQ Y G +L+ ++A  
 Sbjct: 438 NIASSPPITSWFTDKGSRATATQALEYSSNTYMQIAIKRLGQQYVPGMSLSTDNMEKA 497

25 Query: 476 MAKLRKTYGEYGLGVSTGLDLP-ESEGYVPGKYSLGTTLMESFGQYDAYTPMQLGQYIST 534  
 M LR TY E+G+GVSTGLDLP ESEGY+P Y++ L E+FGQYD+YT +QL QY+++  
 Sbjct: 498 MTTLRDTYAEFGMGVSTGLDLPGESEGYIPKNYNVANVLTEAFGQYDSYTTIQLAQYVAS 557

30 Query: 535 IANNGNRLAPHVVSIDIYEGNDSNKFQQLVRSITPKTLNKIAISDQELAI IQEGFYNVVNS 594  
 IAN G R+APH+V IY+ + L ++ + LNK+++ ++L IIQ+GF++VVNS  
 Sbjct: 558 IANGGKRVAPHIVGGIYDAGKNGSLGTLSSVTDTRVLNKLSDSKQLGIIQQGFHDVVNS 617

35 Query: 595 GSGYATGTSMRGNVTTISGKTGTAETFAKNVNGQTVSTYNLNAIAYDTNR---KIAVAVM 651  
 GS ATG +M ++ ISGKTGTAET+A + +G +V+T NLNA+AY T + K+AV +M  
 Sbjct: 618 GSSLATGKAMASSIIPISGKTGTAETATDGSNGSVTTVNLNAVAYATAKDGTKLAVGIM 677

Query: 652 YPHVTTDTTKSHQLVARDMIDQY 674  
 YPH +K+HQ + +++ Y  
 Sbjct: 678 YPHALDWKSKAHQNAVKAIMELY 700

A related GBS gene <SEQ ID 8997> and protein <SEQ ID 8998> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop Possible site: -1 Crend: 8  
 McG: Discrim Score: -12.38  
 GvH: Signal Score (-7.5): -5.9  
 Possible site: 35  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -12.42 threshold: 0.0

45 INTEGRAL Likelihood = -12.42 Transmembrane 23 - 39 ( 18 - 46)  
 PERIPHERAL Likelihood = 4.56 355  
 modified ALOM score: 2.98

50 \*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.5967(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55

The protein has homology with the following sequences in the databases:

50.5/71.3% over 683aa

Streptococcus  
 thermophilus  
 60 GP|1685112| penicillin-binding protein 2b Insert characterized

ORF02276(307 - 2322 of 2643)  
 GP|1685112|gb|AAC44614.1||U58210(17 - 700 of 704) penicillin-binding protein 2b  
 {Streptococcus thermophilus}  
 65 %Match = 38.5

-2496-

%Identity = 50.4 %Similarity = 71.2

Matches = 342 Mismatches = 189 Conservative Sub.s = 141

```

5      108      138      168      198      228      258      288      318
NHGR*NS*LPTTCFRI**KIKPCFRILLR*II*SLYKKFRPSWLEFFIIYNILSVCKKPFLL*YNSSQSFSYKELMLNRKK
      :      :      :      :      :      :      :      :
MTSFWEKNSQKWKWKWRQKRK
      10      20

10     348      378      408      438      468      498      528      558
RYRLTVKKQNASIPRRLNLLFFIIXLLFTVLILRLLEQMIGQQSFYMKKLTALTSTYTVKESKARGQIFDAKGVVLENDE
      :      :      :      :      :      :      :      :
EKRANKPRKPVNISRRVYLFGVVFVFLFLLLFARLTYMQVYNKSFYTKKLEDNSKYTVRIASERGQIFDAKGIALTTNQSS
      30      40      50      60      70      80      90      100

15     588      618      648      678      708      738      768      798
RPTVAFSRGNISSQSIKELANKLSHYITLITEVASSDRAKRDYVLADKANYKKVVESLPDSKRYDKFGNHLAESTVYANA
      :      :      :      :      :      :      :      :
KDVITFTFRSNLVSSDTMKSVARLATLVLTETKVTDRQKREFYLADSANYKRVVNDLPNDKKTDFGNKLAEATIYNNA
      110     120     130     140     150     160     170     180

20     828      858      888      918      948      978      1008     1038
VAAVPVSAINYSEDELKVVALFNQMNAFTPTFGSVKLSTGELSDDIKIKLDADKKELLGISVTSNWHRRKKGTSLSDILGT
      :      :      :      :      :      :      :      :
INAVPDEAVDYSEDELKIVYIYSHMNAVSNFSTVILKTADLTDPQIAIVAQQKELNGIRVAKDWERHTSDSSLSPLIGR
      190     200     210     220     230     240     250     260

25     1068     1098     1128     1158     1188     1218     1248     1278
ISTEKAGLPREEVKKYLKKGYSLNDRVGTSTYLEKQYEDDLQGIQIRKVVNKKGVVSDNITQEGKSGRNLLKLTIDLNY
      :      :      :      :      :      :      :      :
VSSSEAGLPQEDAKDYLLKGYALNDRVGTSTYLEKEYEELQGHVREITVDKEGKVDSDKIIQKGSKGNLLKLTIDLDF
      270     280     290     300     310     320     330     340

30     1308     1338     1368     1398     1428     1452     1482     1512
QNKVESILKQYYSSELSSGRASFSEGMYAIAIEPSTGKVLAMAGLKNHDG--NLVDDSLGTIAKNFTPGSVVKGATLSSG
      :      :      :      :      :      :      :      :
QKGVEDILGQQLSSEISGNKATYSEGMYAIVVMNADTGAVLAMAGQKHEQGAQDFKADALGTITDVFTPGSVVKGATLTAG
      350     360     370     380     390     400     410     420

35     1542     1566     1587     1614     1644     1674     1704     1734
WENKVLRGNEVLYDQ--EIAN---IRSWFT-RGLTPISAAQALEYSSNTYMVQVALRLMGQDYNITGDALTDGRGYQEAAMAK
      :      :      :      :      :      :      :      :
WRSGAIIYGDQVLTQPINIASSPPITSWFTDKGSRATATQALEYSSNTYMVQIAIKRLGQQVVPGMSLSTDNMEKAMTT
      430     440     450     460     470     480     490     500

40     1764     1821     1851     1881     1911     1941     1971
LRKTYGEYGLGVSTGLDLP-ESEGYVPGKYSLGTTLMESFGQYDAYTPMQLGQYISTIANNGNRLAPHVVSDIYEGNDSN
      :      :      :      :      :      :      :
LRDTYAEFGMGVSTGLDLPGESEGYIPKNYNVANVLTEAFGQYDSYTTIQLAQYVASIANGGKRVAPHIVGGIYDAGKNG
      510     520     530     540     550     560     570     580

45     2001     2031     2061     2091     2121     2151     2181     2211
KFAQLVRSITPKTLNKAISDQELAIIEGFFYNVNVNSGSGYATGTSMRGNVTTISGKTGTAETFAKNVNGQTVSTYNLNA
      :      :      :      :      :      :      :
SLGLTSSSTVTRVLNKLSDSKQLGIIQQGFHDVVNSGSSLATGKAMASSIIPISGKTGTAETATDGSNGSVTTVNLNA
      590     600     610     620     630     640     650     660

50     2262     2292     2322     2352     2382     2412     2442
IAYDTNR---KIAVAVMYPHVTDTTKSHQLVARDMIDQYISQFTGQ*ERTFECFTQHQLLN*LTAFAQNYRV*VLKQQHV
      :      :      :      :      :      :      :
VAYATAKDGTKLAVGIMYPHALDWKSKAHQNAVKAIMELYQNTNTH
      670     680     690     700

```

SEQ ID 8998 (GBS292) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 9; MW 103kDa).

GBS292-GST was purified as shown in Figure 211, lane 7.

-2497-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2213

A DNA sequence (GBSx2332) was identified in *S.agalactiae* <SEQ ID 6831> which encodes the amino acid sequence <SEQ ID 6832>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10                   bacterial cytoplasm --- Certainty=0.2644 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15                   >GP:CAB51328 GB:AJ131985 phosphoglyceromutase [Streptococcus pneumoniae]  
                   Identities = 219/230 (95%), Positives = 226/230 (98%)

Query: 1    MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIQAAGIEFDLAFTSVLKR 60  
                   MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLI+ AGI+FD A+TSVLKR  
 20   Sbjct: 1    MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIKEAGIKFDQAYTSVLKR 60

Query: 61    AIKTTNLALEAADQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120  
                   AIKTTNLALEA+DQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL  
 25   Sbjct: 61    AIKTTNLALEASDQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120

Query: 121   PPDMAKDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG 180  
                   PP+M +DDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG  
 30   Sbjct: 121   PPNMDRDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG 180

Query: 181   AHGNSIRALVKHIKQLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLGK 230  
                   AHGNSIRALVKHIK LSDDEIMDVEIPNFPPLVFEFDEKLN+VSEYYLGK  
 30   Sbjct: 181   AHGNSIRALVKHIKGLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLGK 230

35                   A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6833> which encodes the amino acid  
                   sequence <SEQ ID 6834>. Analysis of this protein sequence reveals the following:

Possible site: 61

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40                   bacterial cytoplasm --- Certainty=0.2646 (Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45                   Identities = 206/229 (89%), Positives = 214/229 (92%)

Query: 1    MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIQAAGIEFDLAFTSVLKR 60  
                   MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLI+ AGIEFDLAFTSVL R  
 50   Sbjct: 1    MVKLVFARHGESEWNKANLFTGWADVDLSEKGTQQAIDAGKLIKEAGIEFDLAFTSVL/R 60

Query: 61    AIKTTNLALEAADQLWVPVEKSWRLNERHYGGLTGKNKAEAAEQFGDEQVHIWRRSYDVL 120  
                   AIKTTNLALE A QLWVP EKSRLNERHYG LTGKNKAEAAEQF DEQVHIWRRSYDVL  
 50   Sbjct: 61    AIKTTNLALENAGQLWVPTEKSWRLNERHYGALTGKNKAEAAEQFCDEQVHIWRRSYDVL 120

Query: 121   PPDMAKDDEHSAHTDRRYASLDDSVIPDAENLKVTLERALPFWEDKIAPALKDGNVFG 180  
                   PP MAKDDE+SAH DRRYA LD ++IPDAENLKVTLERAL+P+WE+KIAPAL DGNVFG  
 55   Sbjct: 121   PPAMAKDDEYSAHKDRRYADLDPALIPDAENLKVTLERAMPYWEEKIAPALLDGNVFG 180

Query: 181   AHGNSIRALVKHIKQLSDDEIMDVEIPNFPPLVFEFDEKLNLVSEYYLG 229

-2498-

AHGNSIRALVKHIK LSDDEIMDVEIPNFPPLVFE DEKLN+V EYYLG  
 Sbjct: 181 AHGNSIRALVKHIKGLSDDEIMDVEIPNFPPLVFELDEKLNIVKEYYL 229

SEQ ID 6832 (GBS110) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 8; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 10; MW 53.9kDa).

The GBS110-GST fusion product was purified (Figure 204, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 252A), FACS (Figure 252B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2214

A DNA sequence (GBSx2333) was identified in *S.agalactiae* <SEQ ID 6835> which encodes the amino acid sequence <SEQ ID 6836>. This protein is predicted to be triosephosphate isomerase (tpiA). Analysis of this protein sequence reveals the following:

Possible site: 54  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.37 Transmembrane 36 - 52 ( 36 - 52)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC43268 GB:U07640 triosephosphate isomerase [Lactococcus lactis]  
 Identities = 164/252 (65%), Positives = 202/252 (80%)  
 Query: 1 MSRKPFIA GNWKMNKNPPEAKAFIEAVASKLPSSSELVEAGIAAPALTSTVLEAAKGSEL 60  
 MSRKP IAGNWKMNK EA+AF+EAV + LPSS+ VE+ I APAL L+ + +GSEL  
 Sbjct: 1 MSRKPIIAGNWKMNKTLSEAQAFVEAVKNNLPSSDNVESVIGAPALFIAPMAYLRQGSSEL 60  
 Query: 61 KIAAQNSYFENSGAFTGENSPKVLAEAGTDYVIGHSERRDYFHETDQDINKKAKAIFAN 120  
 K+AA+NSYFEN+GAFTGENSP + ++G +Y++IGHSERR+YFHETD+DINKKAKAIFA  
 Sbjct: 61 KLAAENSYFENAGAFTGENSPAIVDLGIEYIIIGHSERRDYFHETDEDINKKAKAIFAA 120  
 Query: 121 GLTPIICCGESLETYEAGKAVEFVGAQVSAALAGLSEEQVSSLVIAYEPIWAIGTGKSAT 180  
 G TPI+CCGE+LET+EAGK E+V Q+ A LAGL+ EQVS+LVIAYEPIWAIGTGK+AT  
 Sbjct: 121 GATPILCCGETLETFEAGKTAEWVSGQIEAGLAGLTAEQVSNLVIAYEPIWAIGTGKTAT 180  
 Query: 181 QDDAQNMCKAVRDVVAADFGQAVADKVRVQYGGSVKPENVAEYMACPDVDGALVGGASLE 240  
 + A C VR V +G+ V++ VR+QYGGSVKPE + MA ++DGALVGGASLE  
 Sbjct: 181 NEIADETCGVVRSTVEKLYGKEVSEAVRIQYGGSVKPEIEGLMAKENIDGALVGGASLE 240  
 Query: 241 AESFLALLDFVK 252  
 A+SFLALL+ K  
 Sbjct: 241 ADSFLALLEMYK 252

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6837> which encodes the amino acid sequence <SEQ ID 6838>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

-2499-

INTEGRAL Likelihood = -1.81 Transmembrane 36 - 52 ( 36 - 52)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1723(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 220/251 (87%), Positives = 237/251 (93%)

10 Query: 1 MSRKPFIAGNWKMKNPPEAKAFIEAVASKLPSELVEAGIAAPALTSTVLEAAKGSEL 60  
MSRKP IAGNWKMNKNP+EAKAF+EAVASKLPS++LV+ +AAPA+ L T +EAAK S L  
Sbjct: 1 MSRKPIIAGNWKMNKNPQEAKAFVEAVASKLPSTDLVDVAVAAPAVDLVTITTEAAKDSVL 60

15 Query: 61 KIAAQNSYFENSGAFTGENSPKVLAEEMGTDYVVIGHSERRDYFHETDQDINKKAKAIFAN 120  
K+AAQN YFEN+GAFTGE SPKVLAEEMG DYVVIGHSERRDYFHETD+DINKKAKAIFAN  
Sbjct: 61 KVAAQNCYFENTGAFTGETSPKVLAEEMGADYVVIGHSERRDYFHETDEDINKKAKAIFAN 120

20 Query: 121 GLTPIICCGESLETYEAGKAVEFVGAQVSAALAGLSEEQVSSLVIAEPIWAIGTGKSAT 180  
GLTPI+CCGESLETYEAGKAVEFVGAQVSAALAGLS EQV+SLV+AYEPIWAIGTGKSAT  
Sbjct: 121 GLTFIVCCGESLETYEAGKAVEFVGAQVSAALAGLSAEQVASLVLAYEPIWAIGTGKSAT 180

25 Query: 181 QDDAQNMCKAVRDVVAADFGQAVADKVRVQYGGSVKPENVAEYMACPDVDGALVGGASLE 240  
QDDAQNMCKAVRDVVAADFGQ VADKVRVQYGGSVKPEN V+YMACPDVDGALVGGASLE  
Sbjct: 181 QDDAQNMCKAVRDVVAADFGQEVADKVRVQYGGSVKPENVDYMACPDVDGALVGGASLE 240

30 Query: 241 AESFLALLDFV 251  
A+SFLALLDF+  
Sbjct: 241 ADSFLALLDFL 251

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2215

35 A DNA sequence (GBSx2334) was identified in *S. galactiae* <SEQ ID 6839> which encodes the amino acid sequence <SEQ ID 6840>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.3050(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:AAB41198 GB:U75481 elongation factor-Tu [Streptococcus mutans]  
Identities = 44/45 (97%), Positives = 45/45 (99%)

50 Query: 1 MVMPGDNVTIEVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 45  
MVMPGDNVTI+VELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA  
Sbjct: 117 MVMPGDNVTIDVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 161

There is also homology to SEQ ID 1022:

Identities = 44/45 (97%), Positives = 44/45 (97%)

55 Query: 1 MVMPGDNVTIEVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 45  
MVMPGDNVTI VELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA  
Sbjct: 371 MVMPGDNVTINVELIHPIAVEQGTTFSSIREGGRTVGSGIVSEIEA 415

-2500-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 2216

A DNA sequence (GBSx2335) was identified in *S.agalactiae* <SEQ ID 6841> which encodes the amino acid sequence <SEQ ID 6842>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -2.66    Transmembrane    81 - 97 ( 80 - 97)
    INTEGRAL    Likelihood = -2.60    Transmembrane    18 - 34 ( 17 - 34)

----- Final Results -----
          bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2217

A DNA sequence (GBSx2336) was identified in *S.agalactiae* <SEQ ID 6843> which encodes the amino acid sequence <SEQ ID 6844>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0596(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 2218

A DNA sequence (GBSx2337) was identified in *S.agalactiae* <SEQ ID 6845> which encodes the amino acid sequence <SEQ ID 6846>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.3559(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.